INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization

Claude Lemaréchal, François Oustry

N° 3710

Juin 1999

THÈME 4

Rapport de recherche

ISSN 0249-6399

Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization

Claude Lemaréchal, François Oustry

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Numopt

Rapport de recherche n° 3710 — Juin 1999 — 40 pages

Abstract: We show that it is fruitful to dualize the integrality constraints in a combinatorial optimization problem. First, this reproduces the known SDP relaxations of the max-cut and max-stable problems. Then we apply the approach to general combinatorial problems. We show that the resulting duality gap is smaller than with the classical Lagrangian relaxation; we also show that linear constraints need a special treatment.

Key-words: Combinatorial optimization, Lagrangian relaxation, SDP relaxation, duality, quadratic constraints, linear matrix inequalities

(Résumé : tsvp)

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Relaxations semi-définies et dualité lagrangienne avec applications en optimisation combinatoire

Résumé : Nous montrons qu'il est fructueux de dualiser les contraintes d'intégralité d'un problème d'optimisation combinatoire. Tout d'abord, ceci reproduit la relaxation SDP classique pour les problèmes de la coupe maximum et du stable maximum. Nous appliquons ensuite cette approche à des problèmes combinatoires généraux. Nous montrons que le saut de dualité résultant est meilleur que celui de la relaxation lagrangienne classique. Nous montrons également que les contraintes linéaires nécessitent un traitement particulier.

Mots-clé : Optimisation combinatoire, relaxation lagrangienne, relaxation SDP, dualité, contraintes quadratiques, inégalités matricielles linéaires

1 Introduction

The aim of this paper is twofold:

(i) to show that some of the well-known SDP relaxations of combinatorial problems (max-cut, max-stable) are particular instances of a general technique: Lagrangian duality;

(ii) to apply this latter technique to rather arbitrary combinatorial problems, thereby proposing relaxation schemes which turn out to be finer than the classical ones.

A third aspect of this paper is *clarification*: we present Lagrangian and SDP duality in a common framework, as simple as possible, and we demonstrate how to apply the "recipe" of [16].

Lagrangian duality, in various situations (nonconvex, quadratic, SDP), is a central tool used in this paper. For the reader's convenience, its main points are explained in an accessible language, in Sections 2 to 4. These sections could be skipped by a reader already expert with Lagrangian duality and linear matrix inequalities. However, Sections 2.2, 3.2, 4.2 and 4.5 do contain definitely new material and should *not* be skipped. Indeed, they deal with two examples from combinatorial optimization: max-cut and max-stable, which we use to illustrate the course of our development. It is there that are given the results mentioned in (i) above. More precisely, 0-1 constraints can be expressed as $x_i^2 - x_i = 0$; applying Lagrangian duality to them gives birth to an SDP problem; and applying duality again to this problem reproduces the previously known and popular SDP relaxations.

Application of this scheme to these two examples opens the way to rather new dualization schemes for general 0-1 programming problems: here again, just dualize the 0-1 constraints, and dualize the result once more – this is our point (ii) above. Such a technique had already been proposed here and there in the literature, see for examples [10], [21]; but it has been generally overlooked so far by the community. In §5, we develop the technique in some detail, showing in particular that some care should be exerted when treating the coupling constraints $Ax = a$.

2 Recalls on Lagrangian duality

Lagrangian duality will be applied to quadratically constrained problems in §4, which in turn are useful for the combinatorial applications considered in this paper. The present section demonstrates the duality mechanism as a *simple* yet *general* tool, covering in particular conic duality – and in particular SDP duality – as well as duality for nonconvex quadratic programs.

2.1 Basic principles

Consider first an optimization problem put in the form

$$\inf f(x)\,, \quad x \in \mathcal{X}\,, \quad g_j(x) = 0,\ j = 1, \ldots, m\,. \tag{1}$$

We introduce the *Lagrangian*, a function of the primal variable x and of the *dual variable* – or *multipliers* – $u \in \mathbb{R}^m$:

$$\mathcal{X} \times \mathbb{R}^m \ni (x, u) \mapsto L(x, u) := f(x) + \sum_{j=1}^{m} u_j g_j(x) = f(x) + u^\top g(x)\,, \tag{2}$$

if $g \in \mathbb{R}^m$ denotes the vector of constraint-values. The *dual function* is then a function of u alone, defined by

$$\mathbb{R}^m \ni u \mapsto \theta(u) := \inf_{x \in \mathcal{X}} L(x, u)\,, \tag{3}$$

and the *dual problem* is

$$\sup \theta(u)\,, \quad u \in \mathbb{R}^m\,. \tag{4}$$

If some of the constraints in (1) are inequalities $g_j(x) \leqslant 0$, then the corresponding dual variables have a sign-constraint: $u_j \geqslant 0$. The dual function is always concave and upper semi-continuous, *independently* of the data $\mathcal{X}$, f and g. The dual problem is therefore always "easy", insofar as (3) is easy to solve. For a general introduction and motivation of this technique, see for example [9, Chap. XII]; or also [5], [17, Chap. 6] for its application to combinatorial problems.

We will denote by val$(\cdot)$ the optimal value of an optimization problem $(\cdot)$. A crucial connection between (1) and (4) is then the *duality gap* val(1)$-$val(4), which is always a nonnegative number. In fact, the dual problem amounts to solving a certain *convex relaxation* of (1), which can be obtained by applying again duality to (4); see for example [3, §16], [4], [11].

Lagrangian duality is a very versatile technique: in a constrained optimization problem giving birth to the formulation (1), all possible combinations are allowed to select the "hard" constraints – those defining $\mathcal{X}$ – and the "soft" constraints – those defining g and then dualized in L. In view of the above arguments, this selection can then be oriented according to two criteria:

- How easily can the dual function be computed in (3); for this, the function $x \mapsto L(x, u)$ should enjoy some *structure* not appearing directly in (1).
- How efficient is the dual problem in terms of (1); said otherwise: how accurate is the convexification of (1). The duality gap val(1)−val(4) quantifies this accuracy.

Section 5, precisely, will review various dualization schemes when the feasible domain in (1) is described by $Ax = a$, $x \in \{0, 1\}^n$.

A very relevant question, then, is whether the duality gap is 0. As is well known, this question heavily relies on *convexity*, but there are also some technicalities dealing with *compactness*. A crucial object for this is the subset of $\mathbb{R}^m$ made up of all possible constraint-values:

$$G := \{g(x) : x \in \mathcal{X}\} = \{\gamma \in \mathbb{R}^m : \gamma = g(x) \text{ for some } x \in \mathcal{X}\}\,, \tag{5}$$

or rather its convex hull. Along this line, the following preliminary result is useful. Its proof in the present general setting is rather arid and we do not give it. The convex case will suffice for our purpose, and this will be Proposition 3.3 below.

Lemma 2.1 *Suppose there is some $\bar{u} \in \mathbb{R}^m$ such that the dual function θ of* (2), (3) *is finite:* $\inf_x L(x, \bar{u}) > -\infty$. *Then the following two statements are equivalent:*

(i) The convex hull of G defined in (5) *contains 0 as an interior point;*

(ii) $\theta(u) \to -\infty$ for $\|u\| \to +\infty$.

2.2 First example: a dual of Max-Cut

Consider the max-cut problem, written in a form suitable for our purpose:

$$\inf \sum_{i,j=1}^{n} Q_{ij} x_i x_j = x^\top Q x\,, \quad x_i^2 = 1,\ i = 1, \ldots, n\,. \tag{6}$$

Needless to say, the constraints $x_i^2 = 1$ just express that $x_i = \pm 1$. The symmetric matrix Q has nonnegative coefficient, null if (i,j) is not an edge of the graph under study; but our development to come is valid for an arbitrary Q,

The above problem has the form (1), with $\mathcal{X} = \mathbb{R}^n$ and $g_j(x) = x_j^2 - 1$. To apply duality, we form the Lagrangian

$$L(x,u) := x^\top Q x + \sum_{i=1}^{n} u_i (x_i^2 - 1) = x^\top (Q + D(u)) x - e^\top u\,. \tag{7}$$

Here $D(u)$ denotes the diagonal matrix constructed from the vector u, and $e \in \mathbb{R}^n$ is the vector of all ones.

Minimizing this Lagrangian with respect to x (on the whole of $\mathbb{R}^n$!) is a trivial operation: if $Q + D(u)$ is not positive semidefinite, we obtain $-\infty$; otherwise, the best to do is to take $x = 0$. This validates the following result:

Theorem 2.2 *The dual function associated with the maxcut problem* (6), (7) *is*

$$\theta(u) = \begin{cases} -e^\top u & \text{if } Q + D(u) \succcurlyeq 0, \\ -\infty & \text{otherwise}\,. \end{cases}$$

The dual problem $\sup \theta(u)$ *is therefore*

$$\sup -e^\top u\,, \quad \text{subject to } Q + D(u) \succcurlyeq 0\,, \tag{8}$$

which has a nonempty compact set of optimal solutions.

Proof. The expressions of the dual function and problem are clear.

Now observe the following facts:

– taking u big enough gives a positive semidefinite $Q + D(u)$;

– when x_i describes $\mathbb{R}$, the constraint $x_i^2 - 1$ describes $[-1, +\infty[$: the set G of (5) is $[-1, +\infty[^m$ (which is convex by itself) contains 0 as an interior point;
– in (8), we certainly have $Q_{ii} + u_i \geqslant 0$, while $-e^\top u \to -\infty$ if some $u_i \to +\infty$: the dual problem (8) is "sup-compact".

All this illustrates Lemma 2.1, and proves at the same time that (8) has a nonempty compact solution set. □

2.3 Conic duality

Here we extend the previous development to the case where constraints have value in a cone. The corresponding theory is by no means new: it can be found for example in [14, Chap. 8] or [15, Chap. 6]; see also [9, § XII.5.3(a)]. Our aim is rather to demonstrate in simple terms how this duality scheme can be developed, giving birth among others to SDP duality, the subject of § 3.1.

Consider instead of (1):

$$\inf f(x)\,, \quad x \in \mathcal{X}\,, \quad g(x) \in K\,, \tag{9}$$

where K is some subset of $\mathbb{R}^m$. The same dualization scheme as in §2.1 can be used: introduce a "slack variable" $y \in \mathbb{R}^m$ and redefine $\mathcal{X}$ to $\mathcal{X} \times K$: (9) can be rewritten as

$$\inf f(x)\,, \quad x \in \mathcal{X},\ y \in K\,, \quad g(x) = y\,.$$

For $(x, y, u) \in \mathcal{X} \times K \times \mathbb{R}^m$, form the "slackened Lagrangian"

$$L'(x, y, u) = f(x) + u^\top (g(x) - y) = L(x, u) - u^\top y\,;$$

here L is the original Lagrangian (2). Remembering the general relation $\inf_{x,y} = \inf_x \inf_y$, we can obtain the corresponding dual function in two steps, minimizing with respect to $x \in \mathcal{X}$ the value

$$\inf_{y \in K} L'(x, y, u) = L(x, u) + \inf_{y \in K} (-u^\top y)\,.$$

Thus, the dual function associated with (9) is $\theta(u) - \sigma_K(u)$, where

$$\sigma_K(u) := \sup_{y \in K} u^\top y$$

is the so-called *support function* of the set K (note: the support function of a set coincides with that of its closed convex hull; K can therefore be assumed closed and convex).

The case of interest for our purpose is when K is a closed convex *cone* of $\mathbb{R}^m$. Then define the *polar cone* of K by

$$K^\circ := \{u \in \mathbb{R}^m \ : \ u^\top y \leqslant 0 \text{ for all } y \in K\} .$$

It is easy to see that the corresponding support function is in this case

$$\sigma_K(u) = \begin{cases} 0 & \text{if } u \in K^\circ , \\ +\infty & \text{otherwise} \end{cases}$$

(if $u \in K^\circ$, $\sigma_K(u) = u^\top 0$; if $u \notin K^\circ$, there is $\bar{y} \in K$ such that $u^\top \bar{y} > 0$ and then $u^\top(t\bar{y})$ goes to $+\infty$ when $t \to +\infty$).

Thus, when K is a (closed convex) cone, the dual problem of (9) is

$$\sup \theta(u) , \quad u \in K^\circ . \tag{10}$$

With comparison to (1), the introduction of K amounts to chopping off the complement of K° from the dual space. Needless to say, (1) corresponds to $K = \{0\}$ and $\{0\}^\circ = \mathbb{R}^m$.

Remark 2.3 *It can be seen that Lemma 2.1(i) has here the following form: there are points $x^1, \dots, x^{m+1}$ in $\mathcal{X}$ and convex multipliers $\alpha^1, \dots, \alpha^{m+1}$ such that the corresponding convex combination $\sum \alpha^k g(x^k)$ lies in the interior of K. This is called a* Slater *condition, which will be used in Theorem 3.4.* □

3 Recalls on Linear Matrix Inequalities

An instance of (9) which is relevant for our forthcoming development is semidefinite programming, in which the constraint-space is $\mathcal{S}_n$, the space of symmetric $n \times n$ matrices, and $K = \mathcal{S}_n^+$ is the set of positive semidefinite matrices. The latter set is obviously a cone. The resulting *SDP duality* is well developed in [23], where the starting tool is Legendre-Fenchel conjugacy; see also [1]. The present section is closer to the development of A. Shapiro in [19, Chap. 3]. We revisit SDP duality as a particular case of Lagrangian/conic duality, and we have several motivations for this:

– to use a unique theoretical framework, which will apply also to nonconvex quadratic problems;
– to introduce the reader with some elementary facts about linear matrix inequalities, which will also be instrumental for our development;
– to prepare a forthcoming work in a more general framework, where constraints take their values in a space of operators.

3.1 SDP duality

A convenient scalar product on our constraint-space $\mathcal{S}_n$ is the ordinary dot-product in $\mathbb{R}^{n^2}$:

$$(\mathcal{S}_n \times \mathcal{S}_n) \ni (X, U) \mapsto \langle X, U\rangle := \operatorname{tr} XU = \sum_{i,j=1}^{n} X_{ij} U_{ij}\,,$$

also called the trace-product. The reason it is so convenient is the following relation: for all $X \in \mathcal{S}_n$ and $u \in \mathbb{R}^n$,

$$[\textstyle\sum_{i,j=1}^n X_{ij} u_i u_j =]\; u^\top X u = \langle X, uu^\top\rangle\,, \tag{11}$$

which can be checked in a straightforward way (note: $(uu^\top)_{ij} = u_i u_j$); it will be used continually.

Note that $\mathcal{S}_n^+ = \{X \in \mathcal{S}_n : \lambda_n(X) \geqslant 0\}$, where the smallest eigenvalue $\lambda_n(X) = \min_{\|u\|=1} u^\top X u$ appears as a concave continuous function of the matrix X; this shows that the cone $K = \mathcal{S}_n^+$ is convex and closed. For the same reason, its interior is the cone of positive definite matrices. In what follows, we will use the notation $X \succ 0$ [resp. $\succcurlyeq 0$] to express that the (symmetric) matrix X is [resp. semi-]positive definite. Also, the norm of a matrix in $\mathcal{S}_n$ has a useful expression:

$$\|X\|^2 = \operatorname{tr}(X^2) = \sum_{i=1}^{n} \lambda_i(X^2) = \sum_{i=1}^{n} \lambda_i^2(X)\,. \tag{12}$$

The polar cone of $\mathcal{S}_n^+$ with respect to our scalar product is easy to obtain:

Lemma 3.1 *The cone $K = \mathcal{S}_n^+$ is anti-polar: $K^\circ = -K$.*

Proof. $[K^\circ \subset -K]$ Take $X \in K^\circ$ and assume for contradiction that $X \notin -K$: there is $u \in \mathbb{R}^n$ such that $u^\top X u > 0$. Then, in view of (11), $U := uu^\top$ $(\in K!)$ satisfies $\langle X, U\rangle = u^\top X u > 0$, which contradicts $X \in K^\circ$.

$[-K \subset K^\circ]$ Let $X \in -K$, *i.e.*, X is negative semidefinite. Then write the spectral decomposition $X = \sum_{i=1}^n \lambda_i e_i e_i^\top$, where the family $\{e_i\}_{i=1,\dots,n}$ forms an orthonormal basis of $\mathbb{R}^n$ and $\lambda_n \leqslant \ldots \leqslant \lambda_1$ are the non-positive eigenvalues of X. For all $U \in K$, we have

$$\begin{array}{rcll} \langle X, U\rangle & = & \sum \lambda_i \langle U, e_i e_i^\top\rangle & \\ & = & \sum \lambda_i e_i^\top U e_i & [\text{use (11)}] \\ & \leqslant & 0\,. & [\lambda_i \leqslant 0 \text{ and } e_i^\top U e_i \geqslant 0] \end{array} \qquad \square$$

Now, a linear optimization problem over the cone of positive semidefinite matrices (a linear SDP problem) is the following instance of (9):

$$\inf f(x) := b^\top x\,, \quad x \in \mathcal{X} := \mathbb{R}^m\,, \quad g(x) := A_0 + \sum_{j=1}^m x_j A_j \succcurlyeq 0\,, \tag{13}$$

where $b \in \mathbb{R}^m$ is given, as well as each A_j in $\mathcal{S}_n$. Writing the Lagrangian dual (10) of this conic problem is an easy exercise:

– Form the Lagrangian at $(x, U) \in \mathbb{R}^m \times K^\circ = \mathbb{R}^m \times \mathcal{S}_n^-$

$$L(x, U) = b^\top x + \langle g(x), U\rangle = \sum_{j=1}^m (b_j + \langle A_j, U\rangle) x_j + \langle A_0, U\rangle\,, \tag{14}$$

– Calculate the dual function: for $U \succcurlyeq 0$,

$$\theta(U) = \begin{cases} \langle A_0, U\rangle & \text{if } b_j + \langle A_j, U\rangle = 0 \text{ for } j = 1, \dots, m\,, \\ -\infty & \text{otherwise}\,. \end{cases}$$

– Remember that $K^\circ = -K = -\mathcal{S}_n^+$ (Lemma 3.1).
– Eliminate all dual points U at which $\theta(U) = -\infty$ (they play no role in the maximization of θ).
– Altogether, the dual of (13) comes naturally as

$$\sup \langle A_0, U\rangle\,, \quad U \preccurlyeq 0\,, \quad b_j + \langle A_j, U\rangle = 0\,,\ j = 1, \dots, m\,. \tag{15}$$

Another easy exercise is to compute the dual of (15). We form the Lagrangian

$$\mathcal{S}_n^- \times \mathbb{R}^m \ni L'(U,x) := \langle A_0, U\rangle + \sum_{j=1}^m x_j(b_j + \langle A_j, U\rangle)\,,$$

in which we recognize (14), with the same domain of definition. The dual

$$\inf_{x\in\mathbb{R}^m} \sup_{U\preccurlyeq 0} \left(b^\top x + \langle A_0 + \textstyle\sum_{j=1}^m x_j A_j, U\rangle\right)$$

of (15) is thus nothing but (13), which thus appears as its own *bidual*.

3.2 Example: a bidual of Max-Cut

The example of §2.2 can be used again to illustrate our development so far: indeed the dual obtained in (8) is an SDP problem, to which (conic) duality can in turn be applied, using a ("bi")dual variable $X \in \mathcal{S}_n$. Some signs must be changed because (8) is a maximization problem; we take $X \in -K^\circ = \mathcal{S}_n^+$ – see (10) and Lemma 3.1 – and form the ("2nd") Lagrangian

$$L'(u,X) = -e^\top u + \langle X, Q + D(u)\rangle\,, \tag{16}$$

Call $d(X)$ the vector of $\mathbb{R}^n$ constructed from the diagonal of the matrix $X \in \mathcal{S}_n$. Realizing that $\langle X, D(u)\rangle = d(X)^\top u$ (in fact, the operators d and D are adjoint to each other), we have

$$L'(u,X) = (-e + d(X))^\top u + \langle X, Q\rangle\,,$$

whose supremum with respect to u is certainly $+\infty$ if $d(X) - e \neq 0$.

Theorem 3.2 *The dual problem associated with* (8)*,* (16) *is*

$$\inf\,\langle Q, X\rangle\,,\quad d(X) = e\,,\quad X \succcurlyeq 0\,, \tag{17}$$

which has a nonempty compact set of optimal solutions.

Proof. From the discussion above, it is clear enough that the dual function $\theta'(X) = \sup_u L'(x,u)$ is $\langle X, Q\rangle$ if $d(X) = e$, and $+\infty$ if not. The resulting dual problem $\inf_{X\succcurlyeq 0} \theta'(X)$ is just (17).

Now the feasible set in (17) is obviously nonempty; the identity is feasible (and positive definite). Besides, all feasible matrices X have nonnegative eigenvalues which sum up to n (the trace of X); in view of (12), they are bounded. $\square$

Note that this result could be anticipated from Theorem 2.2: (8) satisfies both conditions in Lemma 2.1, hence its dual (17) should satisfy them as well.

The interesting point in the above result is that (17) is just the relaxed problem of [6]. This suggests that the dualization (7), (8) is "gap-equivalent" to the relaxation of Goemans and Williamson. For this, we must check that there is no duality gap between (8) and (17); but this will be a consequence of Theorem 3.4 below.

3.3 Primal-dual equivalence

We now study the relationship between our primal-dual pair of SDP problems (13) and (15). The duality gap will be the object of Theorem 3.4, first we prove Lemma 2.1. Even though we restrict ourselves to SDP duality, we use the notation of §2.1, which is lighter.

Proposition 3.3 *Consider the convex problem* (13) *and its dual* (15)*. Assume that the feasible domain of* (15) *is nonempty. Then the following two properties are equivalent:*

(i) (13) *satisfies the Slater condition: there is $\bar{x} \in \mathbb{R}^m$ with $g(\bar{x})$ positive definite;*

(ii) the objective function $\theta(U)$ in (15) *tends to $-\infty$ when $\|U\| \to +\infty$ while staying feasible.*

Proof. $[(i) \Rightarrow (ii)]$ Take $\bar{x}$ as stated; because positive definite matrices form an open set, there is $\varepsilon > 0$ such that

$$g(\bar{x}) + P \in K \quad \text{for all } P \in \mathcal{S}_n \text{ with } \|P\| \leqslant \varepsilon\,. \tag{18}$$

Then take an arbitrary nonzero $U \in K^\circ$; setting $P_U := \varepsilon U/\|U\|$, we have $g(\bar{x}) + P_U \in K$, therefore

$$0 \geqslant \langle U, g(\bar{x}) + P_U \rangle = \langle U, g(\bar{x}) \rangle + \varepsilon \|U\|\,.$$

Thus,

$$\theta(U) \leqslant L(\bar{x}, U) = f(\bar{x}) + \langle U, g(\bar{x})\rangle \leqslant f(\bar{x}) - \varepsilon\|U\|$$

for all $U \preccurlyeq 0$; this implies (ii).

$[(ii) \Rightarrow (i)]$ For the converse, observe that, when x describes $\mathcal{X} = \mathbb{R}^m$, $g(x)$ describes a convex set (an affine subspace) in $\mathcal{S}_n$. Then "not (i)" means that this set does not meet the interior of K, it is *separated* by some hyperplane in $\mathcal{S}_n$ (see [18, Thm. 11.3] for example). There is a nonzero $\bar{M} \in \mathcal{S}_n$ and a number $\bar{r}$ such that

$$\langle U, \bar{M}\rangle \leqslant \bar{r} \leqslant \langle \bar{M}, g(x)\rangle \quad \text{for all } U \succ 0 \text{ and } x \in \mathbb{R}^m$$

(here we have used the fact that positive definite matrices are dense in $\mathcal{S}_n^+$). Because $0 \in K$, we certainly have $\bar{r} \geqslant 0$; and because K is a cone, we may just assume $\bar{r} = 0$ ($\langle U, \bar{M}\rangle > 0$ would imply $t\langle U, \bar{M}\rangle \to +\infty$ for $t \to +\infty$, contradicting $\langle tU, \bar{M}\rangle \leqslant \bar{r}$). In other words we have

$$\langle U, \bar{M}\rangle \leqslant 0 \leqslant \langle \bar{M}, g(x)\rangle \quad \text{for all } U \in K \text{ and } x \in \mathbb{R}^m\,.$$

It follows in particular that $\bar{M} \in K^\circ$. Now take $\bar{U} \in K^\circ$ such that $\theta(\bar{U}) > -\infty$. We have for all $t \geqslant 0$ and $x \in \mathbb{R}^m$

$$L(x, \bar{U} + t\bar{M}) = L(x, \bar{U}) + t\langle \bar{M}, g(x)\rangle \geqslant L(x, \bar{U}) \geqslant \theta(\bar{U})\,.$$

hence $\theta(\bar{U} + t\bar{M}) \geqslant \theta(\bar{U})$. Observe that $\bar{U} + t\bar{M} \in K^\circ$ and let $t \to +\infty$ to contradict (ii). $\square$

With respect to Lemma 2.1, convexity of G makes the proof a lot easier mainly for the $(i) \Rightarrow (ii)$ part. Without such convexity, points of the form $g(x^k)$ need to be considered (see Remark 2.3), for which the corresponding values $f(x^k)$ need to be controlled... and then it is linearity of f that helps (or at least its boundedness from above).

It is important to understand that the particular form of Slater condition used here is relevant only when a full-dimensional cone K is involved in the constraints. For a problem (1) containing only "natural" equalities, one should return to the general form used in Lemma 2.1(i); incidentally, this form would make easier the proof below. Finally w mention that the concept of interior point is not completely satisfactory (for example, it does not cover a simple linear program where $g(x) = Ax - a$, with a degenerate matrix A); actually, it is the *relative* interior that should be used.

Theorem 3.4 *In the conditions of Proposition 3.3, there is no duality gap: val*(15)=*val*(13).

Proof. Perturb the constraints in (13); the result is the so-called perturbation function

$$\mathcal{S}_n \ni P \mapsto v(P) := \{\inf f(x) : x \in \mathbb{R}^n, g(x) + P \succcurlyeq 0\}.$$

Convexity of the data allows to establish convexity of this function (see [9, § IV.2.4] for example). Besides, the perturbed feasible domain is nonempty for P small enough: see (18). Thus, we have a convex function v, well-defined in a neighborhood of 0: it has a nonempty subdifferential at 0 (see [9, Prop. IV.1.2.1] for example), *i.e.,* there exists $U \in \mathcal{S}_n$ such that

$$v(P) \geqslant v(0) + \langle U, P \rangle \quad \text{for all } P \in \mathcal{S}_n .$$

By definition of v, this means

$$f(x) \geqslant \mathrm{val}(13) + \langle U, P \rangle \quad \text{for all } (x, P) \in \mathbb{R}^n \times \mathcal{S}_n \text{ such that } g(x) + P \succcurlyeq 0 .$$

For given x, the set of P such that $g(x) + P \succcurlyeq 0$ is $-g(x) + K$. Thus, we have

$$f(x) + \langle U, g(x) \rangle \geqslant \mathrm{val}(13) + \langle U, Y \rangle \quad \text{for all } x \in \mathbb{R}^n \text{ and } Y \in K .$$

When Y varies freely in K, the righthand side is bounded from above: U is certainly in K°. Then, taking $Y = 0$, we obtain

$$\mathrm{val}(15) \geqslant \theta(U) = \inf_x f(x) + \langle U, g(x) \rangle \geqslant \mathrm{val}(13) . \qquad \square$$

By symmetry (see the end of §3.1), the above results can be applied to (15) and its dual (13):

Corollary 3.5 *Assume the feasible domain of* (13) *is nonempty. Then there is no duality gap under any of the following equivalent conditions:*

(i) there is $\bar{U} \prec 0$ *feasible in* (15)*;*

(ii) the objective function $b^\top x$ *in* (13) *tends to* $+\infty$ *when* $\|x\|$ *tends to* $+\infty$ *while staying feasible.*

3.4 Infimum of a quadratic function

Technically, the essential content of this paper consists in applying Lagrangian duality to optimization problems involving quadratic functions. As it happens, linear matrix inequalities and their accompanying SDP duality appear naturally in this context, and the aim of this section is to demonstrate the connection. Given $Q \in \mathcal{S}_p$, $b \in \mathbb{R}^p$ and $c \in \mathbb{R}$, consider the function

$$\mathbb{R}^p \ni y \mapsto q(y) := y^\top Q y + b^\top y + c\,.$$

The following result is elementary.

Lemma 3.6 *A necessary and sufficient condition for q to have a finite lower bound over $\mathbb{R}^n$ ($\inf q > -\infty$) is*

$$Q \succcurlyeq 0 \quad \textit{and} \quad b \in \mathcal{R}(Q)\,. \tag{19}$$

Proof. [Necessity] If $Q \not\succcurlyeq 0$, *i.e.,* q is not convex, there exists $y \in \mathbb{R}^p$ such that $y^\top Q y < 0$. Then $q(ty) \to -\infty$ when $t \to +\infty$.

On the other hand, $b \notin \mathcal{R}(Q)$ means that b has a nonzero component $b_{\mathcal{N}}$ on $(\mathcal{R}(Q))^\perp = (\mathcal{R}(Q^\top))^\perp = \mathcal{N}(Q)$. Then

$$q(-tb_{\mathcal{N}}) = b^\top(-tb_{\mathcal{N}}) + c = -t\|b_{\mathcal{N}}\|^2 + c\,,$$

which tends to $-\infty$ when $t \to +\infty$.

[Sufficiency] Conversely, the property $b \in \mathcal{R}(Q)$ means that the linear system $2Qy = -b$ (*i.e.,* $\nabla q(y) = 0$) has at least one solution, and this *characterizes* the minimum points of the convex function q. $\square$

When (19) holds, the minimization of q can be carried out explicitly in terms of the spectral decomposition $Q = \sum_{i=1}^{p-k} \lambda_i q_i q_i^\top$. Here $0 \leqslant k \leqslant p$ is the rank of Q, $\lambda_1 > \ldots > \lambda_{p-k} > 0$ and $q_1, \ldots, q_{p-k}$ are orthonormal (eigen)vectors forming a basis of $\mathcal{R}(Q)$. The *pseudo-inverse* of Q is then

$$Q^\dagger := \sum_{i=1}^{p-k} \lambda_i^{-1} q_i q_i^\top\,.$$

Lemma 3.7 *When* (19) *holds and using the above notation,* $\bar{y} := -\frac{1}{2}Q^\dagger b$ *minimizes* q *and the corresponding infimal value is*

$$\inf_{y\in\mathbb{R}^p} q(y) = q(\bar{y}) = c - \tfrac{1}{4}b^\top Q^\dagger b\,.$$

Proof. We have

$$QQ^\dagger = \sum_{i,j=1}^{p-k} \lambda_i\lambda_j^{-1} q_i(q_i^\top q_j)q_j^\top = \sum_{i=1}^{p-k} q_i q_i^\top\,.$$

Because the q_i's form an orthonormal basis of $\mathcal{R}(Q)$, we can write $b = \sum \beta_i q_i$, with $\beta_i = q_i^\top b$. This gives

$$QQ^\dagger b = \sum_{i,j=1}^{p-k} q_i q_i^\top q_j q_j^\top b = \sum_{i=1}^{p-k} q_i q_i^\top b = b$$

(actually, $QQ^\dagger$ characterizes the orthogonal projection onto $\mathcal{R}(Q)$). Then $Q\bar{y} = -\frac{1}{2}QQ^\dagger b = -\frac{1}{2}b$, *i.e.*, $\bar{y}$ minimizes q. The computation of $q(\bar{y})$ is then straightforward. □

These elementary results can be used to prove a lemma very useful in the analysis of linear matrix inequalities. It is everyday bread in control theory, [2, Chap. 2], and will be useful for us as well. Here we give a simple proof.

Lemma 3.8 (Schur's Lemma) *For* $Q \in \mathcal{S}_p$, $P \in \mathcal{S}_n$ *and* $S = [s_1, \dots, s_p] \in \mathbb{R}^{n\times p}$*, the following two statements are equivalent:*

(i) The matrix $\begin{bmatrix} P & S^\top \\ S & Q \end{bmatrix}$ *is positive [resp. semi-]definite.*

(ii) The matrices Q *and* $P - S^\top Q^\dagger S$ *are both positive [resp. semi-]definite, and* $s_i \in \mathcal{R}(Q)$*, for* $i = 1, \dots, p$*; these last range conditions are automatically satisfied when* $Q \succ 0$ *(then* $\mathcal{R}(Q) = \mathbb{R}^p$*).*

Proof. For $(x, y) \in \mathbb{R}^n \times \mathbb{R}^p$, define the function

$$q_x(y) := \begin{pmatrix} x^\top & y^\top \end{pmatrix} \begin{bmatrix} P & S^\top \\ S & Q \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^\top P x + 2\,(Sx)^\top y + y^\top Q y$$

and consider Lemma 3.6 with $b = 2Sx$, $c = x^\top Px$; observe that (i) means exactly $q_x(y) > 0$ [resp. $\geqslant 0$] for all nonzero $(x, y) \in \mathbb{R}^n \times \mathbb{R}^p$.

[$(i) \Rightarrow (ii)$] Under (i), we always have $Sx \in \mathcal{R}(Q)$.

– Take $x = 0$ to see that (i) implies $Q \succ 0$ [resp. $\succcurlyeq 0$].

– Then take successively $x = e_i$, $i = 1, \dots, p$ so that $Sx = s_i$, which lies in $\mathcal{R}(Q)$.

Finally use Lemma 3.7 to write

$$\inf_{y \in \mathbb{R}^p} q_x(y) = x^\top Px - (Sx)^\top Q^\dagger Sx = x^\top (P - S^\top Q^\dagger S)x\,.$$

This must be positive [resp. nonnegative] for all nonzero $x \in \mathbb{R}^n$, which was the last property we wanted.

[$(ii) \Rightarrow (i)$] Under (ii), we can use again Lemma 3.7 to write for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^p$,

$$q_x(y) \geqslant \inf_{y \in \mathbb{R}^p} q_x(y) = x^\top (P - S^\top Q^\dagger S)x\,.$$

To say that this number is positive [resp. nonnegative] is just (i). $\square$

We give now two applications of Schur's Lemma.

Lemma 3.9 (Rank-one constraints) *The set of $(X, x) \in \mathcal{S}_n \times \mathbb{R}^n$ satisfying $X \succcurlyeq xx^\top$ is closed and convex. Actually,*

$$X - xx^\top \succcurlyeq 0 \ \textit{[resp. } \succ 0\textit{]} \quad \Longleftrightarrow \quad \begin{bmatrix} X & x \\ x^\top & 1 \end{bmatrix} \succcurlyeq 0 \ \textit{[resp. } \succ 0\textit{]}.$$

Proof. The set of (X, x)'s described by the right part of the equivalence is the pre-image of the closed [resp. open] convex set $\mathcal{S}^+_{n+1}$ [resp. $\mathcal{S}^{++}_{n+1}$] by an affine operator; as such, it is closed [resp. open] and convex. To prove the equivalence, apply Lemma 3.8 with $p = 1$, $P = X$, $S = x^\top$ and $Q = 1$. $\square$

Lemma 3.10 (Debreu's Lemma) *Let $A \in \mathbb{R}^{n\times m}$ and $H \in \mathcal{S}_n$ be positive definite on $\mathcal{N}(A)$ ($x^\top Hx > 0$ whenever $Ax = 0$ and $x \neq 0$). Then $H + \pi AA^\top$ is positive definite for all $\pi > 0$ large enough.*

Proof. Calling m' the rank of A, let $Z_{\mathcal{N}}$ and $Z_{\mathcal{R}}$ be respectively $n \times (n - m')$ and $n \times m'$ matrices whose columns form orthonormal bases of $\mathcal{N}(A)$ and $\mathcal{N}(A)^\perp$. The columns of the $n \times n$ matrix $B := [Z_{\mathcal{R}}\, Z_{\mathcal{N}}]$ are orthonormal: B is nonsingular, $B^{-1} = B^\top$ and a matrix M is positive definite if and only if $B^\top M B$ is. Using the block structure of B, we have

$$B^\top M B = \begin{bmatrix} Z_{\mathcal{R}}^\top \\ Z_{\mathcal{N}}^\top \end{bmatrix} M[Z_{\mathcal{R}}\, Z_{\mathcal{N}}] = \begin{bmatrix} Z_{\mathcal{R}}^\top M Z_{\mathcal{R}} & Z_{\mathcal{R}}^\top M Z_{\mathcal{N}} \\ Z_{\mathcal{N}}^\top M Z_{\mathcal{R}} & Z_{\mathcal{N}}^\top M Z_{\mathcal{N}} \end{bmatrix}$$

Since $AZ_{\mathcal{N}} = 0$, we obtain in particular

$$B^\top[H + \pi A^\top A]B = \begin{bmatrix} Z_{\mathcal{R}}^\top H Z_{\mathcal{R}} + \pi Z_{\mathcal{R}}^\top A^\top A Z_{\mathcal{R}} & Z_{\mathcal{R}}^\top H Z_{\mathcal{N}} \\ Z_{\mathcal{N}}^\top H Z_{\mathcal{R}} & Z_{\mathcal{N}}^\top H Z_{\mathcal{N}} \end{bmatrix} =: \begin{bmatrix} P & S^\top \\ S & Q \end{bmatrix}.$$

Now apply Lemma 3.8. Observe that $Q = Z_{\mathcal{N}}^\top H Z_{\mathcal{N}}$ is positive definite (by assumption); it is therefore "onto" and the only thing we need is $P - S^\top Q^\dagger S \succ 0$, *i.e.*,

$$Z_{\mathcal{R}}^\top H Z_{\mathcal{R}} + \pi Z_{\mathcal{R}}^\top A^\top A Z_{\mathcal{R}} \succ Z_{\mathcal{R}}^\top H Z_{\mathcal{N}} (Z_{\mathcal{N}}^\top H Z_{\mathcal{N}})^{-1} Z_{\mathcal{N}}^\top H Z_{\mathcal{R}}\,. \tag{20}$$

We claim that $Z_{\mathcal{R}}^\top A^\top A Z_{\mathcal{R}}$, which obviously lies in $\mathcal{S}_{m'}^+$, is actually positive definite: indeed, for all $v \in \mathbb{R}^{m'}$, the property $v^\top Z_{\mathcal{R}}^\top A^\top A Z_{\mathcal{R}} v = \|AZ_{\mathcal{R}} v\|^2 = 0$ implies

$$Z_{\mathcal{R}} v \in \mathcal{N}(A) \cap \mathcal{R}(A) = \mathcal{N}(A) \cap \mathcal{N}(A)^\perp = \{0\}\,,$$

and hence $v = 0$.

Thus, with simplified notation, we have to prove that a certain matrix $\pi M - N$ is positive definite. Calling R (a symmetric positive definite matrix) the square root of $M = Z_{\mathcal{R}}^\top A^\top A Z_{\mathcal{R}} \succ 0$, we have for all nonzero $v \in \mathbb{R}^{m'}$:

$$v^\top(\pi M - N)v = \pi v^\top R R v - v^\top N v = \pi \|Rv\|^2 - v^\top R R^{-1} N R^{-1} R v\,.$$

Setting $w = Rv$ $(\neq 0)$, this is equal to $\pi\|w\|^2 - w^\top R^{-1} N R^{-1} w$, which is a positive number for π large enough, namely for $\pi > \lambda_1(R^{-1}NR^{-1})$. $\square$

4 Lagrangian duality for quadratic programs

The material introduced in Sections 2 and 3 is just what is needed to apply Lagrange duality to quadratic optimization problems.

4.1 Dualizing a quadratic problem

Let be given $m+1$ quadratic functions defined on $\mathbb{R}^n$:

$$q_j(x) := x^\top Q_j x + b_j{}^\top x + c_j\,,\ j = 0, \ldots, m\,,$$

where the Q_j's lie in $\mathcal{S}_n$, the b_j's in $\mathbb{R}^n$ and the c_j's in $\mathbb{R}$; we assume $c_0 = 0$. With these data, consider the quadratic problem

$$\left\{\begin{array}{l} \inf q_0(x)\,, \quad x \in \mathbb{R}^n \\ q_j(x) = 0,\ j = 1, \ldots, m\,. \end{array}\right. \tag{21}$$

The associated Lagrangian (2) of §2.1 with $\mathcal{X} := \mathbb{R}^n$, $f := q_0$ and $g_j := q_j$, $j = 1, \ldots, m$ is then

$$L(x,u) = x^\top Q(u) x + b(u)^\top x + c(u)\,, \tag{22}$$

where $Q(u) := Q_0 + \sum_{j=1}^m u_j Q_j$, $b(u) := b_0 + \sum_{j=1}^m u_j b_j$ and $c(u) = \sum_{j=1}^m u_j c_j = c^\top u$. Just as in §2.1, we denote by

$$\mathbb{R}^m \ni u \mapsto \theta(u) := \inf_{x \in \mathbb{R}^n} L(x,u) \tag{23}$$

the dual function, and the dual problem is to maximize θ. In fact the dual function can be computed explicitly:

Proposition 4.1

$$\theta(u) = \left\{\begin{array}{ll} c(u) - \frac{1}{4} b(u)^\top Q(u)^\dagger b(u) & \textit{if } Q(u) \succcurlyeq 0 \textit{ and } b(u) \in \mathcal{R}(Q(u))\,, \\ -\infty & \textit{otherwise.} \end{array}\right.$$

Proof. Just apply Lemma 3.6 and its extension Lemma 3.7. □

Then Schur's Lemma enables a "PD-representation" of the epigraph of the dual function (23) (terminology appearing in [15, § 6.4.3]), and thereby an SDP-formulation of the dual problem:

Corollary 4.2 *The dual of* (21), (22) *is equivalent to the SDP problem with variables $u \in \mathbb{R}^m$ and $r \in \mathbb{R}$:*

$$\left\{\begin{array}{l} \sup r\,, \\ \left[\begin{array}{cc} c(u) - r & \frac{1}{2} b(u)^\top \\ \frac{1}{2} b(u) & Q(u) \end{array}\right] \succcurlyeq 0\,. \end{array}\right. \tag{24}$$

In other words, $\sup_{u\in\mathbb{R}^m}\theta(u)$ *is the optimal value* $\bar r$ *of* (24) *(a number in* $[-\infty,+\infty]$*).*

Proof. Using an extra variable $r \in \mathbb{R}$, the dual problem consists in maximizing r subject to $\theta(u) \geqslant r$. In view of Proposition 4.1, this last property exactly means

$$c(u) - r - \frac{1}{4}b(u)^\top Q(u)^\dagger b(u) \geqslant 0\,, \quad Q(u) \succcurlyeq 0\,, \quad b(u) \in \mathcal{R}(Q(u))$$

and from Lemma 3.8, this is equivalent to the SDP constraint in (24). □

Thus, even though the dual problem is posed in $\mathbb{R}^m$, it involves intermediate variables in $\mathcal{S}_n$, and even in $\mathcal{S}_{n+1}$. This is somehow made necessary by the requirement $Q(u) \succcurlyeq 0$, unavoidable in (22), (23).

4.2 Second example: Max-Stable

We have applied Lagrangian duality to the max-cut problem to illustrate how this general strategy finds back a known and successful particular relaxation technique. The same phenomenon occurs with another problem having a popular relaxation: the maximum stable problem, which we formulate as follows:

$$\max w^\top x\,, \quad x_i x_j = 0,\ (i,j) \in E\,, \quad x_i^2 = x_i,\ i = 1,\dots,n\,; \tag{25}$$

here we have used again the trick $t^2 - t = 0 \Leftrightarrow t \in \{0,1\}$. Of course $w > 0$ (if some $w_i \leqslant 0$, just set the corresponding x_i to 0! typically $w = e$); but this assumption is useless.

We use the dualization technique of §4.1. The notation is slightly complicated, though. Of course we change signs because we maximize; but also, there are now two sets of dual variables:

(i) those assigned to the constraints $x_i x_j = 0$; it is convenient to consider in (25) only those constraints with $i < j$ (E contains no (i,i)!) and to call $-\frac{1}{2}U_{ij}$ the corresponding multiplier; then the contribution to the Lagrangian of these terms will be

$$-\frac{1}{2}\sum_{\substack{(i,j)\in E\\ i<j}} U_{ij}x_i x_j = -\sum_{(i,j)\in E} U_{ij}x_i x_j\,;$$

(ii) those for the constraints $x_i^2 - x_i = 0$, as in §2.2; we call $-U_{ii}$ these multipliers, which produce the contribution $-\sum U_{ii}x_i^2 + \sum U_{ii}x_i$.

Thus, the dual variables altogether form a symmetric matrix U, in which we impose $U_{ij} = 0$ if the constraint (i,j) is not present in (25). In summary, we take the Lagrangian

$$L(x,U) = w^\top x - \sum_{i,j=1}^{n} U_{ij}x_i x_j + \sum_{i=1}^{n} U_{ii}x_i = -x^\top U x + (d(U)+w)^\top x\,, \tag{26}$$

whith the dual domain

$$\mathcal{U} := \{U \in \mathcal{S}_n \ : \ U_{ij} = 0 \text{ if } \ i \neq j \text{ and } (i,j) \notin E\}\,.$$

Maximizing this Lagrangian (with respect to x unconstrained) can be done with the tools of §4.1, yielding an SDP representation of the dual epigraph.

Theorem 4.3 *The dual value* $\inf \theta(U)$ *associated with* (25), (26) *is the optimal value of*

$$\left\{\begin{array}{l} \inf r\,, \quad (U,r) \in \mathcal{U} \times \mathbb{R}\,, \\ \begin{bmatrix} r & -\frac{1}{2}(d(U)+w)^\top \\ -\frac{1}{2}(d(U)+w) & U \end{bmatrix} \succcurlyeq 0\,. \end{array}\right. \tag{27}$$

Proof. The function $-L(x,U)$ of (26) is an instance of (22), with adapted notation and sign. Then the dual problem (24) from §4.1 becomes (27). □

Slater and compactness properties could be checked on this problem, but it will be simpler to use the bidual for this.

4.3 Bidualizing a quadratic problem

Now, duality can again be applied to the SDP problem (24). This bidualization of the starting primal problem (21) is actually a *convexification* of it.

Theorem 4.4 *The dual of* (24), *i.e., the bidual of* (21), *is*

$$(SDP) \qquad \left\{\begin{array}{l} \inf \langle Q_0, X\rangle + b_0^\top x\,, \quad X \in \mathcal{S}_n,\ x \in \mathbb{R}^n\,, \\ \langle Q_j, X\rangle + b_j^\top x + c_j = 0,\ j = 1,\dots,m\,, \\ \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succcurlyeq 0\,. \end{array}\right.$$

Proof. The result comes mechanically from SDP duality (§ 3.1). One can also detail the calculations, according to § 2.3 in the cone $K = \mathcal{S}^+_{n+1}$. We take the (bi)dual variable

$$Y = \begin{bmatrix} t & x^\top \\ x & X \end{bmatrix} \in -K^\circ = \mathcal{S}^+_{n+1}$$

(see Lemma 3.1) and form the Lagrangian associated with (24)

$$\begin{array}{rcl} L(u,r,Y) & = & r + \left\langle \begin{bmatrix} t & x^\top \\ x & X \end{bmatrix}, \begin{bmatrix} c(u) - r & \frac{1}{2}b(u)^\top \\ \frac{1}{2}b(u) & Q(u) \end{bmatrix} \right\rangle \\ & = & (1-t)r + tc(u) + x^\top b(u) + \langle X, Q(u) \rangle \\ & = & (1-t)r + \langle X, Q_0 \rangle + x^\top b_0 + \sum_{j=1}^m u_j(\langle X, Q_j \rangle + x^\top b_j + tc_j)\,. \end{array}$$

Its supremum with respect to (u, r) (the (bi)dual function, to be minimized) is $+\infty$ unless $t = 1$ and the coefficient of each u_j is zero. Then the only remaining term is $\langle X, Q_0 \rangle + x^\top b_0$. Altogether, we recognize (SDP). □

Another exercise is left to the reader: applying once more SDP duality to (SDP), one obtains (24) again.

Observe that the bidual (SDP) of (21) appears as a *convex* problem; this is normal, a dual problem is always convex (and (24) is convex as well). Applying weak duality twice, we see that val(21)$\geqslant$val(24)$\leqslant$val(SDP). Comparing val(21) with val(SDP) requires one more argument, which we give now.

4.4 The lifting procedure

Looking at (SDP) does reveal a strong similarity with (21); basically, these two problems just differ by the introduction of the new variable X. The similarity is made even more blatant with the help of Lemma 3.9: (SDP) can also be written

$$\left\{ \begin{array}{l} \inf \langle Q_0, X \rangle + b_0^\top x\,, \quad X \in \mathcal{S}_n,\ x \in \mathbb{R}^n\,, \\ \langle Q_j, X \rangle + b_j^\top x + c_j = 0,\ j = 1, \ldots, m\,; \\ X \succcurlyeq xx^\top\,. \end{array} \right. \tag{28}$$

In addition, this writing allows a direct interpretation of (SDP), without passing by the intermediate problem (24):

- In view of (11), a quadratic form $x^\top Q x$ can equivalently be written as $\langle Q, xx^\top \rangle$.
- Setting $X := xx^\top$, (21) can therefore be written as

$$\begin{cases} \inf \langle Q_0, X\rangle + \langle b_0, x\rangle \,, \quad X \in \mathcal{S}_n, x \in \mathbb{R}^n \,, \\ \langle Q_j, X\rangle + \langle b_j, x\rangle + c_j = 0, \; j = 1, \ldots, m \,, \\ X = xx^\top \,. \end{cases}$$

- In this last problem, everything is linear except the last constraint, which is nonconvex.
- *Relax* this last constraint $X = xx^\top$ to $X \succcurlyeq xx^\top$, which is now convex with respect to (x, X) (Lemma 3.9). Needless to say, this widens the feasible domain.

The convexification mechanism is now clear. The above "recipe" is called *lifting procedure* in [13], where it was introduced in the context of 0-1 programming. This procedure thus appears as fairly general, indeed: it is applicable to any optimization problem involving quadratic functions; but any polynomial can be reduced to a quadratic form, via appropriate additions of variables (see for example [21], [22]). As a result, the lifting procedure can be considered as a convexification process for polynomial optimization.

Let us summarize the relative merits of the convex relaxations introduced in this section, in terms of duality gaps:

Proposition 4.5 *There holds*

$$val(21) \geqslant val(28) = val(SDP) \geqslant val(24).$$

Proof. The last inequality is weak duality between (24) and its dual (SDP) (beware of the change of signs!) The equality is Lemma 3.9. For the first inequality, take any x feasible in (21) (if there is none, val(21)$=+\infty$ and there is nothing to prove); then $(x, X := xx^\top)$ is feasible in (28) and has the same objective value $q_0(x)$ (use (11), in particular). Thus, the feasible domain in (28) is "larger" than in (21). □

It is worth mentioning that the last inequality in Proposition 4.5 often holds as an equality, namely in every "normal" situation where there is no duality gap between the pair of dual problems (24) and (SDP) (see Theorem 3.4).

Remark 4.6 *We emphasize the difference between the two dualization schemes resulting in the lifting procedure:*

– *One* – (21)→(24) – *dualizes m constraints, producing a dual problem in $\mathbb{R}^m$;*

– *the second* – (24)→(28) – *dualizes $O(n^2)$ constraints, producing* (28) *with $O(n^2)$ variables.*

These two dualization schemes have little to do with each other. As a result, (28) *is an "exotic bidual" of* (21)*; both problems are not even posed in the same space.* □

4.5 Examples: Max-Stable and Max-Cut

Just as Lagrangian duality reproduced Goemans-Williamson relaxation for the max-cut problem, it turns out to reproduce for the max-cut problem one of the forms of Lovász's number [12] (see [8]). Such a result was already mentioned in [22]. Before proving it, we state separately a result which will be also used later.

Lemma 4.7 *The set of matrices $X \in \mathcal{S}_n$ satisfying $X \succcurlyeq d(X)d(X)^\top$ is compact.*

Proof. First, this set is obviously closed. The assumption certainly implies that $X \succcurlyeq 0$, hence $X_{ii} \geqslant 0$, and also that the diagonal of $X - d(X)d(X)^\top$ is nonnegative: $X_{ii} - X_{ii}^2 \geqslant 0$; altogether, $X_{ii} \leqslant 1$. It follows that $\operatorname{tr} X \leqslant n$, each eigenvalue of X is bounded (by n). In view of (12), X is bounded. □

Theorem 4.8 *Problem* (27) *and*

$$\left\{\begin{array}{l} \sup w^\top d(X)\,, \quad X \in \mathcal{S}_n\,, \\ X_{ij} = 0,\ (i,j) \in E\,, \\ X - d(X)d(X)^\top \succcurlyeq 0 \end{array}\right. \tag{29}$$

have the same optimal value. Furthermore, both problems have a nonempty compact set of optimal solutions.

Proof. Write (25) as an instance of (21): the objective function is $q_0(x) := -w^\top x$; the constraints are $q_{ij}(x) := x_i x_j$ and $q_i(x) := x_i^2 - x_i$. Then apply Theorem 4.4: the dual of (27) – the bidual of (25) – is

$$\left\{\begin{array}{l} \sup w^\top x \,, \\ X_{ij} = 0, \ (i,j) \in E \,, \\ X_{ii} - x_i = 0, \ i = 1, \dots, n \,. \\ \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succcurlyeq 0 \,. \end{array}\right. \tag{30}$$

This implies $d(X) = x$ and, using Lemma 3.9, we obtain (29).

Now take the matrix $X = \frac{1}{2}\mathrm{I}$, so that $X - d(X)d(X)^\top = \frac{1}{4}\mathrm{I}$; it satisfies Slater's condition in (29); besides, the feasible domain in (29) is compact (Lemma 4.7). Altogether, there is no duality gap between the pair of dual problems (27) and (29), and both have a nonempty compact set of optimal solutions (Theorem 3.4). □

The lifting procedure of (4.4) could have been applied to (25), thus producing (29) directly. This procedure could also be applied to (6), we would obtain the relaxation

$$\inf \langle Q, X \rangle \,, \quad d(X) = e \,, \quad X \succcurlyeq xx^\top \,.$$

However the variable x is superfluous: the best to do is to take $x = 0$, which gives (17) indeed.

5 Application to 0–1 programming

Consider now a rather general combinatorial optimization problem:

$$\inf b^\top x \,, \quad Ax = a \,, \quad x_i^2 - x_i = 0, \ i = 1, \dots, n \,. \tag{31}$$

By contrast with the examples seen so far, nothing is quadratic except the Boolean constraints; but linear (equality) constraints do exist. As a result, even without assuming any structure in the matrix A, this problem has at least three possible dualizations following the pattern of §2.1 or 4.1.

5.1 Dualization L

The particular choice $g = Ax - a \in \mathbb{R}^m$, $\mathcal{X} = \{0,1\}^n$ produces the Lagrangian

$$L_{\mathrm{L}}(x,u) := b^\top x + u^\top (Ax - a)\,.$$

Its minimization over $\mathcal{X} = \{0,1\}^n$ is trivial, resulting in a piecewise linear dual function:

$$\theta_{\mathrm{L}}(u) := \sum_{i=1}^{n} \min\{0, (b + A^\top u)_i\} - a^\top u\,. \tag{32}$$

Maximizing θ_{L} is therefore a nonsmooth optimization problem, which can be done by a number of methods, say those of [7], of [20], or of Chaps XII and XV in [9]. As far as duality gaps are concerned, the resulting dual problem amounts to replacing $\{0,1\}^n = \mathcal{X}$ in (31) by its convex hull $[0,1]^n$.

Remark 5.1 *This last property can for example be obtained by an easy exercise: form an LP equivalent to the problem* $\sup \theta_L$ *and dualize it. As illustrated by the previous sections, duality generates a convexification, which can be characterized by a bi-dualization. It is interesting to note that the present convexification is performed in the primal space* $\mathbb{R}^n$ itself*; remember Remark 4.6.* □

This scheme is quite popular. It can even be said that the expression "Lagrangian relaxation" of (31) in the combinatorial community tacitly means the present dualization L_{L} (possibly moving some of the constraints $Ax = a$ into the definition of $\mathcal{X}$). We will call it dualization L(inear).

5.2 Dualization C

Another possible dualization of (31) consists in taking $\mathcal{X} = \mathbb{R}^n$ and in dualizing all the constraints: the linear as well as the Boolean ones. This produces another Lagrangian, which now depends on two dual variables $(u,v) \in \mathbb{R}^m \times \mathbb{R}^n$:

$$\begin{aligned} L_{\mathrm{C}}(x,u,v) &= b^\top x + u^\top (Ax - a) + \textstyle\sum_{i=1}^n v_i (x_i^2 - x_i) \\ &= (b + A^\top u - v)^\top x + x^\top D(v) x - a^\top u\,. \end{aligned}$$

Computing the associated dual function $\theta_{\mathrm{C}}(u,v)$ can be done With the tools of §4.1, or via elementary calculations: $\theta_{\mathrm{C}}(u,v) = -a^\top u + \sum_{i=1}^n \theta_i(u,v_i)$, where

$$\theta_i(u,v_i) := \begin{cases} -\frac{(b+A^\top u - v)_i^2}{4v_i} & \text{if } v_i > 0\,, \\ 0 & \text{if } v_i = 0 \text{ and } (b+A^\top u)_i = 0\,, \\ -\infty & \text{otherwise}\,. \end{cases} \tag{33}$$

Theorem 5.2 *The dualizations L and C produce the same duality gap. More precisely:*

$$\theta_{\mathrm{L}}(u) = \sup_{v\in\mathbb{R}^n} \theta_{\mathrm{C}}(u,v)\,. \tag{34}$$

Proof. Clearly, $\sup_v \theta_{\mathrm{C}}(u,v) = -a^\top u + \sum_i \sup_{v_i} \theta_i(u,v_i)$. Now the maximization of each $\theta_i(u,\cdot)$ is an easy exercise illustrating §4.3: setting $\alpha := (b+A^\top u)_i$, we obtain 0 for $\alpha = 0$, while for $\alpha \neq 0$ (which forces $v_i > 0$):

$$\sup_{t\in\mathbb{R}} \theta_i(u,t) = \sup_{t>0} \theta_i(u,t) = \frac{1}{2}\alpha - \frac{1}{2}|\alpha| = \min\{0,\alpha\}\,.$$

Comparing with (32), this proves (34). Then the C-dual value is

$$\sup_{u,v} \theta_{\mathrm{C}}(u,v) = \sup_u \Big(\sup_v \theta_{\mathrm{C}}(u,v)\Big) = \sup_u \theta_{\mathrm{L}}(u)\,,$$

which is just the L-dual value. □

Remark 5.3 *The real reason underlying this result is worth meditating. In fact, $\theta_L(u)$ is the optimal value of the problem*

$$\inf L_{\mathrm{L}}(x,u)\,, \quad x_i^2 = x_i,\ i = 1,\dots,n\,.$$

Suppose we solve it via duality: we form a Lagrangian, which is nothing else than $L_{\mathrm{C}}(x,u,v)$. For lack of compactness, the minimization of $L_{\mathrm{C}}(\cdot,u,v)$ requires $v \geqslant 0$; said otherwise, the constraints $x_i^2 = x_i$ can equivalently be taken as inequalities $x_i^2 \leqslant x_i$. The present dual approach ignores the nonconvexity *of the original problem. Since, in addition, a Slater condition holds (when x_i describes $\mathbb{R}$, the constraint $x_i^2 - x_i$ describes $[-\frac{1}{4}, +\infty[$, which contains 0 as interior point), there is no duality gap:* (34) *holds.* □

This dualization, which we will call C(omplete), is therefore of moderate interest as compared to L of §5.1: it increases the number of dual variables, it produces a more complicated dual function, all this without improving the duality gap. Note also that maximizing θ_{L} can somehow be seen as maximizing θ_{C}, with priority given to v. By contrast, dualization B of §5.3 below gives priority to u; but with a significant refinement, though.

Remark 5.4 *Applying the lifting procedure of §4.4 to* (31)*, we obtain*

$$\begin{cases} \inf b^\top d(X)\,, \quad X \in \mathcal{S}_n\,, \\ Ad(X) = a\,, \\ X - d(X)d(X)^\top \succcurlyeq 0\,. \end{cases}$$

Despite the appearances, this is an ordinary LP in $\mathbb{R}^n$. Actually, X plays no role except through its diagonal: we can set $X_{ij} = 0$ for $i \neq j$ and replace $d(X)$ by $x \in \mathbb{R}^n$. The SDP constraint becomes $x - xx^\top \succcurlyeq 0$, which just means $0 \leqslant x \leqslant 1$. □

5.3 Dualization B

A third possibility for (31) is to put the linear constraints in $\mathcal{X}$ and dualize the Boolean constraints. In other words, we do just the opposite to L: we take the Lagrangian

$$L_{\mathrm{B}}(x, v) := (b - v)^\top x + x^\top D(v)x \tag{35}$$

and the dual function is then the optimal value of the problem

$$\theta_{\mathrm{B}}(v) = \inf_{Ax=a} x^\top D(v)x + (b - v)^\top x\,. \tag{36}$$

We will call B(oolean) this dualization scheme. Computing θ_{B} is still possible with the tools of §4. However, before developing the calculations, we check that this complication is worth the effort.

Theorem 5.5 *There holds* $\sup \theta_{\mathrm{B}} \geqslant \sup \theta_{\mathrm{C}} = \sup \theta_{\mathrm{L}}$.

Proof. The equality is Theorem 5.2. The inequality is a general result: when more constraints are dualized, the duality gap can only increase (see for example [11]). □

Actually, a more careful comparison of dualizations B and C explains why B improves the duality gap:

Lemma 5.6 *If* $v_i \geqslant 0$, $i = 1, \ldots, n$, *then* $\theta_{\mathrm{B}}(v) = \sup_u \theta_{\mathrm{C}}(u, v)$. *As a result, the* C*-dual value is given by*

$$\sup_{u,v} \theta_{\mathrm{C}}(u, v) = \sup_{v \in \mathbb{R}^n_+} \theta_{\mathrm{B}}(v) \,.$$

Proof. When $v \in \mathbb{R}^n_+$, the B-Lagrangian problem (36) is a standard linear-quadratic program, for which there is no duality gap:

$$\theta_{\mathrm{B}}(v) = \sup_{u \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^n} [L_{\mathrm{B}}(x, v) + u^\top (Ax - a)] = \sup_{u \in \mathbb{R}^m} \theta_{\mathrm{C}}(u, v) \,.$$

Now, $\theta_{\mathrm{C}}(u, v) \equiv -\infty$ if $v \notin \mathbb{R}^n_+$ (see (33)): to compute the C-dual value, we may as well insert the constraint $v \geqslant 0$:

$$\sup_{u,v} \theta_{\mathrm{C}}(u, v) = \sup \{\theta_{\mathrm{C}}(u, v) : u \in \mathbb{R}^m, v \geqslant 0\} = \sup_{v \geqslant 0} \theta_{\mathrm{B}}(v) \,. \qquad \square$$

Thus, dualizations C or L consist in maximizing θ_{B} over $\mathbb{R}^n_+$. By contrast, dualization B maximizes θ_{B} over a *possibly larger* set. In particular, θ_{B} assumes finite values on

$$\check{V} := \{v \in \mathbb{R}^n : D(v) \succ 0 \text{ on } \mathcal{N}(A)\} \,. \tag{37}$$

Even more: for $v \in \check{V}$, (36) has a strictly convex objective function, and therefore a unique optimal solution $x(v)$, which varies continuously with v. This has important consequences:

Proposition 5.7 *With the above notation, assume:*

(i) θ_{C} *has some maximum point* $(\bar{u}, \bar{v})$ *with* $\bar{v} \in \check{V}$,

(ii) the C-duality gap is nonzero.

then the B-duality gap is strictly smaller than the L-duality gap.

Proof. Use the continuity on $\breve{V}$ of the unique solution $x(v)$: for v in a neighborhood of $\bar{v}$ (ensuring in particular $v \in \breve{V}$), some extra constraint $\|x - x(\bar{v})\| \leqslant \varepsilon$ can be inserted in (36) without changing its solution $x(v)$. Then the so-called filling property XII.2.3.1 of [9] holds: θ_{B} is differentiable in this neighborhood, with derivatives $x_i^2(v) - x_i(v)$, $i = 1, \dots, n$.

Now use Lemma 5.6: if dualization B does not improve the duality gap, $\sup \theta_{\mathrm{B}} = \theta_{\mathrm{B}}(\bar{v})$. This implies $\nabla\theta_{\mathrm{B}}(\bar{v}) = 0$: $x(\bar{v})$ is feasible in (31). Since $x(\bar{v})$ minimizes $L_{\mathrm{C}}(\cdot, \bar{u}, \bar{v})$, we have a C-primal-dual solution: dualization C – hence L – has no duality gap. $\square$

Assumption (i) means that the domain of θ_{B} contains a useful point, not lying in the v-part of the domain of θ_{C}. This assumption is guaranteed to hold for example when $\breve{V} \supset \mathbb{R}^n_+$. Mutatis mutandis, the efficiency of dualization B will increase when $\breve{V}$ increases, i.e. when $\mathcal{N}(A)$ decreases, i.e. when there are more constraints.

We now turn to actual computation of θ_{B}. The case $a \notin \mathcal{R}(A)$ (i.e. $Ax = a$ has no solution in $\mathbb{R}^n$) is uninteresting: then val(31) is $+\infty$ and $\theta_{\mathrm{B}} \equiv +\infty$; there is no "duality gap". We therefore assume that $a = A\hat{x}$ for some $\hat{x}$ ($\in \mathbb{R}^n$). Besides, calling m' the rank of A, let Z be a matrix whose $n - m'$ columns form a basis of $\mathcal{N}(A)$. Thus:

$$Ax = a \;\text{ with } x \in \mathbb{R}^n \quad \Longleftrightarrow \quad x = \hat{x} + Zy \;\text{ with } y \in \mathbb{R}^{n-m'},$$

so that (31) is equivalent to the problem in y

$$\left\{\begin{array}{l} \inf c^\top Zy + b^\top \hat{x}\,, \\ y^\top Z_i^\top Z_i y + (2\hat{x}_i - 1)Z_i y + \hat{x}_i^2 - \hat{x}_i = 0\,,\; i = 1, \dots, n\,. \end{array}\right. \tag{38}$$

Here $Z_i \in \mathbb{R}^{n-m'}$ denotes the ith row of Z.

Theorem 5.8 *The dual value* $\sup_v \theta_{\mathrm{B}}(v)$ *associated with* (31), (35) *is the value of the SDP problem*

$$\left\{\begin{array}{l} \sup r\,, \\ \begin{bmatrix} \hat{c}(v) - r & \frac{1}{2}\hat{b}(v)^\top \\ \frac{1}{2}\hat{b}(v) & Q(v) \end{bmatrix} \succcurlyeq 0\,, \end{array}\right.$$

where

$$\begin{aligned} \hat{c}(v) &:= x^\top D(v)\hat{x} + (b-v)^\top \hat{x}\,, \\ \hat{b}(v) &:= 2D(v)\hat{x} + b - v\,, \\ Q(v) &:= ZD(v)Z^\top\,. \end{aligned}$$

Proof. Plug the change of variable $x = \hat{x} + Z^\top y$ in (36) and use the above notation $\hat{c}(v)$, $\hat{b}(v)$ and $Q(v)$:

$$\theta_{\mathrm{B}}(v) = \inf_{y \in E} y^\top Q(v) y + \hat{b}(v)^\top y + \hat{c}(v)\,.$$

Not unexpectedly, we recognize in the infimand the Lagrangian associated with (38): we are in the situation of §4.1 and the above SDP problem is just the corresponding instance of (24). □

Naturally, a convexified form of (31) will be obtained as before by bi-dualization. However, the computations turn out to be lighter if a simplified but equivalent form of dualization B is used; we now explain it.

5.4 A simpified form of B

As suggested in [16], another dualization is possible, which avoids the heavy algebra implied by Theorem 5.8, without worsening the B-duality gap. The idea is to solve (36) by the so-called *augmented Lagrangian* technique: add a penalty term $\pi\|Ax - a\|^2$ (which is 0 for x feasible!) to the objective function L_{B} and use duality. In other words, taking $\pi \geqslant 0$ and setting

$$L^\pi(x, u, v) := L_{\mathrm{B}}(x, v) + \pi\|Ax - a\|^2 + u^\top(Ax - a)\,, \tag{39}$$

replace (36) by the problem

$$\sup_{u \in \mathbb{R}^m} \theta^\pi(u, v)\,, \quad \text{where} \quad \theta^\pi(u, v) := \inf_{x \in \mathbb{R}^n} L^\pi(x, u, v)\,.$$

The following result demonstrates the validity of the idea (cf. Lemma 5.6).

Lemma 5.9 *If $v \in \check{V}$ of (37), then $\theta_{\mathrm{B}}(v) = \sup_u \theta^\pi(u, v)$ for π large enough. It follows that*

$$\theta_{\mathrm{B}}(v) = \sup_{(u,\pi) \in \mathbb{R}^{m+1}} \theta^\pi(u, v) \quad \textit{if } v \in \check{V}\,.$$

Proof. Clearly, the additional term $\pi\|Ax-a\|^2$ in the infimand of (36) does not change the optimal value $\theta_{\mathrm{B}}(v)$. Now invoke Debreu's Lemma 3.10: if $v \in \check{V}$ then, for π large enough, the penalized (36) has a positive definite Hessian $D(v)+\pi A^\top A$. This problem becomes an ordinary linear-quadratic program, to which we can apply duality theory:

$$\theta_{\mathrm{B}}(v) = \sup_{u\in\mathbb{R}^m} \inf_{x\in\mathbb{R}^n} L_{\mathrm{B}}(x,v) + u^\top(Ax-a) + \pi\|Ax-a\|^2$$

and our first claim is proved.

Then observe that the function $\pi \mapsto L^\pi(x,u,v)$ is nondecreasing. This property is transmitted to the infimum in x ($\theta^\pi(u,v)$ increases with π), and then to the supremum in u; our second claim is proved. $\square$

The equivalence between the present dualization and B will be proved if the restriction $v \in \check{V}$ can be removed in the above lemma. Such is the case indeed if, at the same time, we maximize with respect to v – but this is precisely what we want.

Theorem 5.10 *There holds*

$$\sup_{v\in\mathbb{R}^n} \theta_{\mathrm{B}}(v) = \sup\left\{\theta^\pi(u,v) : u\in\mathbb{R}^m, v\in\mathbb{R}^n, \pi \geqslant 0\right\}.$$

Proof. The key argument is the fact that the infimum of a convex function f is not changed if restricted to the relative interior of its domain:

$$[-f^*(0) =] \quad \inf_{v\in\mathbb{R}^n} f(v) = \inf_{v\in \operatorname{ri}\operatorname{dom} f} f(v)\,. \tag{40}$$

This is because, if a function f is changed while leaving its closed convex hull f^{**} intact, then the conjugate f^* is not changed; see [9, Cor. X.1.3.6] for example.

All we have to do is therefore show that $\check{V} \supset \operatorname{ri}\operatorname{dom}\theta_{\mathrm{B}}$: in view of Lemma 5.9, the B-dual optimal value will then be

$$\sup_{v\in\mathbb{R}^n} \theta_B(v) = \sup_{v\in\check{V}} \theta_{\mathrm{B}}(v) = \sup_{v\in\check{V}} \left(\sup_{u,\pi} \theta^\pi(u,v)\right).$$

Clearly from Lemma 3.6,

$$\check{V} \subset \operatorname{dom}\theta_{\mathrm{B}} \subset \bar{V}\,, \quad \text{where} \quad \bar{V} := \{v\in\mathbb{R}^n : D(v) \succcurlyeq 0 \text{ on } \mathcal{N}(A)\}\,.$$

Incidentally, $\operatorname{dom}\theta_{\mathrm{B}}$ is full-dimensional (it contains $\mathbb{R}^m_{++}$); and, with the notation of Theorem 5.8, we have the equivalent definitions

$$\check{V} = \{v \in \mathbb{R}^n : ZD(v)Z^\top \succ 0\}\,, \quad \bar{V} = \{v \in \mathbb{R}^n : ZD(v)Z^\top \succcurlyeq 0\}\,.$$

Since the eigenvalues of $ZD(v)Z^\top$ are continuous in v, it is clear that $\check{V}$ and $\bar{V}$ are the interior and closure of each other. The required property is true. $\square$

The dual and bidual problems are rather simple to write: the approach amounts to writing (31) as

$$\begin{cases} \inf b^\top x\,, \quad Ax = a\,, \quad x_i^2 - x_i = 0,\ i = 1, \dots, n\,, \\ \|Ax - a\|^2 = 0\,. \end{cases}$$

Assigning a multiplier π to the redundant constraint and forming the C-Lagrangian, we obtain nothing else than $L^\pi(x, u, v)$ of (39). The resulting dual problem is then given by §4.1: $\sup_v \theta_{\mathrm{B}}(v) = \sup_{u,v,\pi} \theta^\pi(u, v)$ is the optimal value of

$$\begin{cases} \sup r\,, \\ \begin{bmatrix} \pi\|a\|^2 - r & \frac{1}{2}(b + A^\top(u - 2\pi a) - v)^\top \\ \frac{1}{2}(b + A^\top(u - 2\pi a) - v) & D(v) + \pi A^\top A \end{bmatrix} \succcurlyeq 0\,. \end{cases}$$

As for the bidual, it is given by §4.3 or the recipe of §4.4:

$$\begin{cases} \inf b^\top d(X)\,, \quad X \in \mathcal{S}_n\,, \\ Ad(X) = a\,, \\ X - d(X)d(X)^\top \succcurlyeq 0\,, \\ \langle A^\top A, X\rangle = \|a\|^2\,. \end{cases} \tag{41}$$

Remark 5.11 *As alreday mentioned, the present §5.4 has its roots in [16]. Actually, it was proposed there to avoid dualization B by mere penalty, replacing* (36) *by*

$$\lim_{\pi \to +\infty} \inf_{x \in \mathbb{R}^n} L_{\mathrm{B}}(x, v) + \pi\|Ax - a\|^2\,.$$

The reason we prefer augmented Lagrangian is that there are serious technical difficulties to prove that simple penalty does not worsen the B-duality gap. $\square$

Let us summarize our development so far. Dualization of (31) can be done in two ways:

– One gives birth to the classical LP relaxation

$$\min b^\top x\,,\quad Ax = a\,,\quad x_i \in [0,1]\,.$$

– Viewed with the SDP glasses of dualization C, this can be written

$$\min b^\top d(X)\,,\quad Ad(X) = a\,,\quad X \succcurlyeq d(X)d(X)^\top\,.$$

Indeed, the matrix X in this last formulation plays a role only through its diagonal; the extra-diagonal terms can be eliminated, which reproduces the LP formulation; see Remark 5.3.

– The second relaxation starts from the previous formulation and appends to it the extra constraint $\langle A^\top A, X\rangle = \|a\|^2$. This gives a chance of improving the gap, in a way described by Proposition 5.7.

Appending redundant constraints is actually a classical idea in nonconvex duality. This is the basis of [13] (append the constraints $x_i(Ax - a) =$ and $(1 - x_i)(Ax - a) = 0$). Here, we can also observe that $x_i x_j \in [0,1]$, hence the constraints $X_{ij} \in [0,1]$ can be appended to the B-relaxation (41).

One more comment: needless to say, dualization B accepts quadratic objectives and/or constraints in (31). This was already the case of max-cut and max-stable. These two problems had no linear constraints, hence there was no ambiguity: only dualization B could be applied

5.5 Inequality constraints; the knapsack problem

Suppose that (31) contains inequality constraints; say

$$\inf b^\top x\,,\quad Ax \leqslant a\,,\quad x_i \in \{0,1\},\ i = 1,\ldots,n\,, \tag{42}$$

or even a mixture of equalities and inequalities.

Dualization L is not really affected by this new situation. The only change is that the dual problem now has positivity constraints and becomes $\sup_{u\geqslant 0} \theta_{\mathrm{L}}(u)$; but the duality gap still corresponds to replacing $\{0,1\}$ by $[0,1]$.

Dualization C is also unaffected. The u-part of the dual variables is again constrained to $\mathbb{R}^m_+$, and we again obtain a scheme equivalent to L:

Theorem 5.12 *For* (42)*, dualizations L and C produce the same duality gap.*

Proof. The statement and the proof of Theorem 5.2 remain valid under the constraint $u \geqslant 0$. □

By contrast, dualization B breaks down: the dual function becomes

$$\inf_{Ax \leqslant a} x^\top D(v)x + (b-v)^\top x\,.$$

Since $D(v)$ has no reason to be positive semi-definite, the above problem is nonconvex and NP-hard.

The conclusion is: dualization B can treat equality constraints only; as for inequalities, they have to be treated by a C strategy. However, the case $m=1$ is somewhat special. Indeed, consider the knapsack problem

$$\min b^\top x\,, \quad \ell^\top x \leqslant a\,, \quad x_i^2 - x_i = 0,\ i = 1,\dots,n\,. \tag{43}$$

The three possible dualizations are here, for $u \geqslant 0$ and $v \in \mathbb{R}^n$:

$$\begin{array}{rcll} L_{\mathrm{L}}(x,u) &=& (b+u\ell)^\top x - au & \text{for } x \in \{0,1\}^n\,, \\ L_{\mathrm{C}}(x,u,v) &=& x^\top D(v)x + (b+u\ell-v)^\top x - au & \text{for } x \in \mathbb{R}^n\,, \\ L_{\mathrm{B}}(x,v) &=& x^\top D(v)x + (b-v)^\top x & \text{for } \ell^\top x \leqslant a\,. \end{array}$$

With its single inequality constraint, B becomes a constructive dualization:

Proposition 5.13 *Suppose $\ell \neq 0$. Then the B-dual function $\theta_{\mathrm{B}}(v)$ of* (43) *is $-\infty$ if $v \notin \mathbb{R}^n_+$.*

Proof. If some $v_i < 0$, call e_i the corresponding canonical basis vector; let $\bar{x}$ be feasible in (43) ($\bar{x}$ exists). Depending on the sign of ℓ_i, $\bar{x} \pm te_i$ remains feasible for $t \to +\infty$, while $L_{\mathrm{B}}(\bar{x} \pm te_i, v) \to -\infty$. □

Thus, the computation of $\theta_{\mathrm{B}}(v)$ has to be done only when $v \geqslant 0$, and then it reduces to an easy linear-quadratic problem. A consequence of this property is that the versatility of Lagrangian dualization is disappears:

Theorem 5.14 *For* (43)*, the dualization B and C produce the same duality gap. More precisely:*

$$\theta_{\mathrm{B}}(v) = \sup_{u \geqslant 0} \theta_{\mathrm{C}}(u,v)\,.$$

As a result, the three dualizations L, C and B produce the same duality gap.

Proof. In view of Proposition 5.13, we may assume $v \geqslant 0$ (otherwise we have the trivial relation $-\infty = \sup_{u\geqslant 0} -\infty$). Then we can apply standard duality theory to the linear-quadratic problem $\inf_{\ell^\top x \leqslant a} L_{\mathrm{B}}(x,v)$:

$$\theta_{\mathrm{B}}(v) = \sup_{u\geqslant 0} \inf_{x\in\mathbb{R}^n} [L_{\mathrm{B}}(x,v) + u(\ell^\top x - a)] = \sup_{u\geqslant 0} \inf_{x\in\mathbb{R}^n} L_{\mathrm{C}}(x,u,v)$$

and our first claim is proved.

Then the second claim comes with Theorem 5.12. □

We mention that the same results hold if (43) contains also linear equality constraints $Ax = a$: just reduce the space to $\mathcal{N}(A)$.

We conclude this section with an illustrative example. Mutatis mutandis, the B-duality gap has chances to be smaller when there are more (equality) constraints. From this point of view, a problem with only one constraint is the least favourable. Indeed consider a problem in $\mathbb{R}^2$:

$$\min c^\top x\,, \quad x_2 = 2x_1 \quad x_i \in \{0,1\}, i = 1,2\,.$$

Eliminating x_2, the feasible x_1's are those satisfying $x_1^2 = x_1$ and $2x_1^2 = x_1$.

The B-lifting procedure consists in

– introducing the extra variable $X \in \mathcal{S}_1 = \mathbb{R}$,

– "linearizing" the quadratic constraints to $X = x_1$, $2X = x_1$, and

– introducing the conic constraint $X \geqslant x_1^2$.

The only feasible point is clearly $X = x_1 = 0$ $(= x_2)$: the relaxed feasible domain is not larger than the original one, the B-gap is 0 (in fact, it can be proved that the B-gap is 0 for $n \leqslant 0$).

By contrast, the feasible domain for the C=L=LP relaxation is the segment with endpoints 0 and $(1/2, 1)$, which may introduce a duality gap depending on the position of the vector c in $\mathbb{R}^2$.

6 Conclusions

Relaxation (i.e. enlargement) of the feasible domain is a central tool in combinatorial optimization. SDP relaxation, which embeds the primal space $\mathbb{R}^n$ into the space of symmetric matrices $\mathcal{S}(\mathbb{R}^n)$, appears to be a possible technique; it has demonstrated its efficiency on some applications.

The motivation for this paper was first to show that this technique actually has its roots in a general tool: Lagrangian duality – which itself should be understood in a broader sense than usual in combinatorial optimization. Among other things, the way is thus open to SDP relaxation as a *general* tool, applicable to a large class of combinatorial problems.

Besides, we have also shown that SDP relaxation can often be more accurate than the classical Lagrangian relaxation. Our analysis suggests the way to construct an SDP relaxation, so as to keep duality gaps as small as possible. For example, we are currently studying its application to problems such as graph coloring, or quadratic assignment.

Finally, it should be mentioned that Lagrangian duality corresponds to a convexification in the primal space *itself*, instead of squaring the number of variables, as is done in SDP relaxation. An interesting question is therefore to study this narrower convexification: comparing the two should give fruitful interpretations for either one. This is currently under study.

A limitation of our approach is that it is clumsy in situation particularly favourable to classical Lagrangian relaxation: when the original problem (31) contains two groups of constraints, say

$$\min b^\top x\,, \quad A_1 x = a_1\,, \quad A_2 x = a_2\,, \quad x \in \{0,1\}^n\,.$$

If the constraints indexed by 1 [resp. 2] are "simple" [resp. "complicating"], a standard technique is to dualize the complicating constraints, forming the dual function

$$\min_x\,(b + A_2^\top u)^\top x - u^\top a_2\,, \quad A_1 x = a_1\,, \quad x \in \{0,1\}^n\,.$$

This may very well result in a better duality gap than the B-dualization. We believe that SDP relaxation might be better suited to problems where there are two groups of *variables*, instead of constraints; but more work is necessary to invent appropriate schemes.

References

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.

[2] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[3] I. Ekeland and R. Temam. *Analyse Convexe et Problèmes Variationnels*. Dunod, Paris, 1974.

[4] S. Feltenmark and K. C. Kiwiel. Dual applications of proximal bundle methods, including lagrangian relaxation of nonconvex problems. *SIAM Journal on Optimization*. to appear.

[5] A.M. Geoffrion. Lagrangean relaxation for integer programming. *Mathematical Programming Study*, 2:82–114, 1974.

[6] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 6:1115–1145, 1995.

[7] J.L. Goffin, A. Haurie, and J.Ph. Vial. Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38(2):284–302, 1992.

[8] C. Helmberg, S. Poljak, F. Rendl, and H. Wolkowicz. Combining semidefinite and polyhedral relaxations for integer programs. In E. Balas and J. Clausen, editors, *Integer Programming and Combinatorial Optimization IV*, volume 920 of *Lecture Notes in Computer Science*, pages 124–134. Springer Verlag, Berlin, 1995.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Number 305-306 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1993.

[10] F. Körner and C. Richter. Zur effektiven lösung von booleschen, quadratischen optimierungsproblemen. *Numerische Mathematik*, 40:99–109, 1982.

[11] C. Lemaréchal and A. Renaud. Dual-equivalent convex and nonconvex problems. *Mathematical Programming*. Submitted.

[12] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979.

[13] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

[14] D.G. Luenberger. *Optimization by Vector Space Methods*. Wiley, New York, 1969.

[15] Yu.E. Nesterov and A.S. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Number 13 in SIAM Studies in Applied Mathematics. SIAM, Philadelphia, 1994.

[16] S. Poljak, F. Rendl, and H. Wolkowicz. A recipe for semidefinite relaxation for (0,1)-quadratic programming. *Journal of Global Optimization*, 7:51–73, 1995.

[17] C.R. Reeves. *Modern Heuristic Techniques for Combinatorial Problems*. Blackwell Scientific Publications, New York, 1993.

[18] R.T. Rockafellar. *Convex Analysis*. Number 28 in Princeton Mathematics Ser. Princeton University Press, 1970.

[19] R. Saigal, L. Vandenberghe, and H. Wolkowicz. *Handbook of Semidefinite-Programming*. Kluwer, 1999.

[20] N.Z. Shor. *Minimization methods for non-differentiable functions*. Springer-Verlag, Berlin, 1985.

[21] N.Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987.

[22] N.Z. Shor. Dual estimates in multiextremal problems. *Journal of Global Optimization*, 2:411–418, 1992.

[23] L. Vandenberghe and S. Boyd. Semidefinite programming. *Siam Review*, 38(1):49–95, 1996.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399