Relaxations and Randomized methods for nonconvex QCQPs

Alexandre d'Aspremont, Stephen Boyd
EE392o, Stanford University

Autumn, 2003

1 Introduction

In EE 364, Convex optimization, we have seen that what makes a problem computationally easy or hard is not linearity or nonlinearity but rather convexity or nonconvexity. In particular, EE 364 detailed an extremely wide range of convex problems which could be solved with very good complexity bounds by an appropriate use of Newton's method. In this light, it is tempting to consider nonconvexity, hence complexity, as only the result of poor problem formulation.

However, nonconvex constraints and objectives *do arise* in important practical problems, often due to some natural limitations such as for example discretization, fixed transaction costs, binary communications, etc...

The question is how can we use some of the efficient methods designed for convex problems to get at least *approximate* solutions or bounds for nonconvex ones? We will detail some of the most successful techniques answering that question. Here, we first focus on Lagrangian relaxations, *i.e.*, using weak duality and the convexity of duals to get bounds on the optimal value of nonconvex problems. In a second section, we show how randomization techniques provide near optimal feasible points with, in some cases, bounds on their suboptimality.

1.1 Problem definition

In this note, we will focus on a specific class of problems: Quadratically Constrained Quadratic Programs, or QCQP (see also §4.4 in [?]). We will see that the range of problems that can be formulated as QCQP is extremely vast (in fact, Shor shows that all polynomial problems can be reduced to QCQP) and we will focus on some specific examples throughout the notes. We write a generic QCQP as:

$$\begin{array}{ll} \text{minimize} & x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & x^T P_i x + q_i^T x + r_i \leq 0, \quad \text{i=1,\dots,m,} \end{array} \qquad (1)$$

in the variable $x \in \mathbf{R}^n$ with parameters $P_i \in \mathbf{S}^n$, $q_i \in \mathbf{R}^n$ and $r_i \in \mathbf{R}$. In the case where all the matrices P_i are positive semidefinite, the problem is convex and can be solved efficiently. Here we will focus on the case where at least one of the P_i is not positive semidefinite. Note that the formulation above implicitly includes problems with equality constraints as an equality constraint is equivalent to two opposing inequalities.

1.2 Examples and applications

We list here some examples of nonconvex problems with important practical applications.

1.2.1 Boolean least squares

The problem is stated as:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array} \tag{2}$$

in the variable $x \in \mathbf{R}^n$. This is a basic problem in digital communications (maximum likelihood estimation for digital signals). As the problem is NP-Hard, a brute force solution is to check for all 2^n possible values of x... But the problem can also be directly written as a QCQP:

$$\begin{array}{ll} \text{minimize} & x^T A^T A x - 2b^T A x + b^T b \\ \text{subject to} & x_i^2 - 1 = 0, \quad i = 1, \ldots, n, \end{array}$$

which is of the form (1).

1.2.2 Minimum cardinality problems

The problem consists in finding a minimum cardinality solution to a set of linear inequalities and can be stated as:

$$\begin{array}{ll} \text{minimize} & \mathbf{Card}(x) \\ \text{subject to} & Ax \preceq b, \end{array} \tag{3}$$

in the variable $x \in \mathbf{R}^n$, with $\mathbf{Card}(x)$ the cardinal of the set $\{i : x_i \neq 0\}$. For simplicity, we assume that the feasible set $\{x : Ax \preceq b\}$ is compact and included in some unit ball with radius $R > 0$. We can then reformulate this problem as a QCQP of the form (1):

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T v \\ \text{subject to} & Ax \preceq b \\ & -Rv \preceq x \preceq Rv \\ & v_i^2 - v_i = 0, \quad i = 1, \ldots, n, \end{array}$$

in the variables $x \in \mathbf{R}^n$ and $v \in \mathbf{R}^n$. This problem has many application in engineering and finance, including for example low-order controller design and portfolio optimization with fixed transaction costs.

1.2.3 Partitioning problems

We consider here the two-way partitioning problem described in §5.1.4 and exercise 5.39 of [**?**]:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array} \tag{4}$$

with variable $x \in \mathbf{R}^n$, where $W \in \mathbf{S}^n$ satisfies $W_{ii} = 0$. This problem is directly a QCQP of the form (1). A feasible x corresponds to the partition

$$\{1, \ldots, n\} \; = \; \{i \mid x_i = -1\} \, \cup \, \{i \mid x_i = 1\},$$

and the matrix coefficient W_{ij} can be interpreted as the cost of having the elements i and j in the same partition, with $-W_{ij}$ the cost of having i and j in different partitions. The objective in (4) is the total cost, over all pairs of elements, and problem (4) seeks to find the partition with least total cost.

1.2.4 MAXCUT

MAXCUT is a classic problem in network optimization and a particular case of the partitioning problem above. Here $W \in \mathbf{S}^n$ is a matrix with positive coefficients describing the network topology, with $W(i,j) = 0$ if no arc connects nodes i and j in the network. The problem is then formulated as:

$$\begin{array}{ll} \text{maximize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array} \tag{5}$$

with variable $x \in \mathbf{R}^n$. The objective here is to find a partition of the set so that the sum of the coefficients $W(i,j)$ of the nodes linking the two partitions is maximized (hence the name MAX CUT).

1.2.5 Polynomial problems

A polynomial problem seeks to minimize a polynomial over a set defined by polynomial inequalities:

$$\begin{array}{ll} \text{minimize} & p_0(x) \\ \text{subject to} & p_i(x) \leq 0, \quad i = 1, \ldots, m. \end{array}$$

While seemingly much more general than simple QCQPs, all polynomial problems can be turned into nonconvex QCQPs. Let us briefly detail how. First, we notice that we can reduce the maximum degree of an equation by adding variables. For example, we can turn the constraint

$$y^{2n} + (\ldots) \leq 0$$

into

$$\begin{array}{l} u^n + (\ldots) \leq 0 \\ u = y^2. \end{array}$$

We have reduced the maximum degree of the original inequality by introducing a new variable and a quadratic equality constraint. We can also get rid of product terms, this time

$$xyz + (\dots) \leq 0$$

becomes

$$\begin{aligned} &ux + (\dots) \leq 0 \\ &u = yz. \end{aligned}$$

Here, we have replaced a product of three variables by a product of two variables (quadratic) plus an additional quadratic equality constraint. By applying these transformations iteratively, we can transform the original polynomials into quadratic objective and constraints, thus turning the original polynomial problem into a QCQP on a larger set of variables.

Example. Let's work out a specific example. Suppose that we want to solve the following polynomial problem:

$$\begin{array}{ll} \text{minimize} & x^3 - 2xyz + y + 2 \\ \text{subject to} & x^2 + y^2 + z^2 - 1 = 0, \end{array}$$

in the variables $x, y, z \in \mathbf{R}$. We introduce two new variables $u, v \in \mathbf{R}$ with

$$u = x^2, \quad v = yz.$$

The problem then becomes:

$$\begin{array}{ll} \text{minimize} & xu - 2xv + y + 2 \\ \text{subject to} & x^2 + y^2 + z^2 - 1 = 0 \\ & u - x^2 = 0 \\ & v - yz = 0, \end{array}$$

which is a QCQP of the form (1), in the variables $x, y, z, u, v \in \mathbf{R}$.

2 Convex relaxations

In this section, we begin by describing some direct relaxations of (1) using semidefinite programming. We then detail how Lagrangian duality can be used as an "automatic" procedure to get lower bounds on the optimal value of the nonconvex QCQP described in (1). Note that both techniques provide lower bounds on the optimal value of the problem but give only a minimal hint on how to find an approximate solution (or even a feasible point...), this will be the object of the next section.

2.1 Semidefinite relaxations

Starting from the original generic QCQP:

$$\begin{array}{ll} \text{minimize} & x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & x^T P_i x + q_i^T x + r_i \leq 0, \quad \text{i=1,\dots,m}, \end{array}$$

using $x^T P x = \mathbf{Tr}(P(xx^T))$, we can rewrite it:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\ \text{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \\ & X = xx^T. \end{array} \tag{6}$$

We can directly relax this problem into a convex problem by replacing the last nonconvex equality constraint $X = xx^T$ with a (convex) positive semidefiniteness constraint $X - xx^T \succeq 0$. We then get a lower bound on the optimal value of (1) by solving the following SDP:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\ \text{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \\ & X \succeq xx^T. \end{array}$$

where the last constraint $X \succeq xx^T$ is convex and can be formulated as a Schur complement (see §A.5.5 in [?]):

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\ \text{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \\ & \begin{bmatrix} X & x^T \\ x & 1 \end{bmatrix} \succeq 0, \end{array}$$

2.2 Lagrangian relaxations

We now study a more rigorous method to get relaxations of nonconvex problems, taking advantage of the fact that the dual of a problem is always convex, hence efficiently solvable. Again, starting from the original problem in (1):

$$\begin{array}{ll} \text{minimize} & x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & x^T P_i x + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \end{array}$$

in the variable $x \in \mathbf{R}^n$ with parameters $P_i \in \mathbf{S}^n$, $q_i \in \mathbf{R}^n$ and $r_i \in \mathbf{R}$. We form the Lagrangian:

$$L(x, \lambda) = x^T \left(P_0 + \sum_{i=1}^m \lambda_i P_i \right) x + \left(q_0 + \sum_{i=1}^m \lambda_i q_i \right)^T x + r_0 + \sum_{i=1}^m \lambda_i r_i$$

in the variables $x \in \mathbf{R}^n$ and $\lambda \in \mathbf{R}^m_+$. In the QCQP case, the dual can be computed explicitly using the fact that (see example 4.5 in [?]):

$$\inf_{x \in \mathbf{R}} x^T P x + q^T x + r = \begin{cases} r - \frac{1}{4} q^T P^\dagger q, & \text{if } P \succeq 0 \text{ and } q \in \mathcal{R}(P) \\ -\infty, & \text{otherwise.} \end{cases}$$

where we have noted $P^\dagger$ the pseudo-inverse of P. We now have:

$$\inf_{x \in \mathbf{R}^n} L(x, \lambda) = -\frac{1}{4} \left(q_0 + \sum_{i=1}^m \lambda_i q_i \right)^T \left(P_0 + \sum_{i=1}^m \lambda_i P_i \right)^\dagger \left(q_0 + \sum_{i=1}^m \lambda_i q_i \right) + \sum_{i=1}^m \lambda_i r_i + r_0$$

and we can form the dual of (1), using Schur complements (cf. §A.5.5):

$$
\begin{array}{ll}
\text{maximize} & \gamma + \sum_{i=1}^{m} \lambda_i r_i + r_0 \\
\text{subject to} & \left[\begin{array}{cc} \left(P_0 + \sum_{i=1}^{m} \lambda_i P_i\right) & \left(q_0 + \sum_{i=1}^{m} \lambda_i q_i\right)/2 \\ \left(q_0 + \sum_{i=1}^{m} \lambda_i q_i\right)^T/2 & -\gamma \end{array} \right] \succeq 0 \\
& \lambda_i \geq 0, \quad i = 1, \ldots, m,
\end{array} \tag{7}
$$

in the variable $\lambda \in \mathbf{R}^m$. As the dual to (1), this is a convex program, it is in fact a semidefinite program. Weak duality implies that its optimum value is a lower bound on the optimal value of (1). Using semidefinite duality (see example 5.1 in [**?**]), we can compute the dual of this last problem, *i.e.* the bidual of (1):

$$
\begin{array}{ll}
\text{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\
\text{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \\
& \left[\begin{array}{cc} X & x^T \\ x & 1 \end{array} \right] \succeq 0,
\end{array} \tag{8}
$$

in the variables $X \in \mathbf{S}^n$ and $x \in \mathbf{R}^n$. We observe that this also corresponds to a relaxation of (6), having changed the constraint $X = xx^T$ into $X \succeq xx^T$ (by Schur complement, see §A.5.5 in [**?**]), hence we have recovered via Lagrangian duality techniques the relaxation found at the beginning of the section. This technique of taking the dual twice produces a lower bound for the generic QCQP in (1).

It is important to keep in mind however that this second dual is not unique. Different choices of domains for x will sometimes produce different duals and different second duals. To complicate matters a little further, some of these choices of dual produce better bounds than others (see [**?**] for a discussion)... We will see how this works out in practice by computing relaxations for some of the example problems described above.

2.3 Examples

Let us now compute the Lagrangian relaxations of the examples detailed above.

2.3.1 MINCARD relaxation

Let's first consider the MINCARD problem detailed in (3):

$$
\begin{array}{ll}
\text{minimize} & \mathbf{Card}(x) \\
\text{subject to} & Ax \preceq b.
\end{array}
$$

If we assume that the feasible set $\{x : Ax \preceq b\}$ is compact and included in some centered ball with radius R, we can reformulate this problem as:

$$
\begin{array}{ll}
\text{minimize} & \mathbf{1}^T v \\
\text{subject to} & Ax \preceq b \\
& -Rv \preceq x \preceq Rv \\
& v \in \{0,1\}^n,
\end{array} \tag{9}
$$

in the variables $x, v \in \mathbf{R}^n$, and we then turn this into a QCQP by replacing the constraints $v_i \in \{0,1\}$ by $v_i^2 - v_i = 0$. The problem then becomes:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T v \\ \text{subject to} & Ax \preceq b \\ & -Rv \preceq x \preceq Rv \\ & v_i^2 - v_i = 0, \quad i = 1, \ldots, n. \end{array}$$

The relaxation given by (8) is then:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T v \\ \text{subject to} & Ax \preceq b \\ & -Rv \preceq x \preceq Rv \\ & \mathbf{Tr}(e_i e_i^T X) - e_i^T x = 0, \quad i = 1, \ldots, n \\ & \left[\begin{array}{cc} X & x^T \\ x & 1 \end{array}\right] \succeq 0, \end{array}$$

where e_i is the Euclidean basis in $\mathbf{R}^n$. Both [**?**] and [**?**, Th. 5.2] show that this relaxation produces the same lower bound as the direct linear programming relaxation:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T v \\ \text{subject to} & Ax \preceq b \\ & -Rv \preceq x \preceq Rv \\ & v \in [0,1]^n, \end{array} \tag{10}$$

which is also, up to a multiplicative constant, the classical l_1 heuristic described in [**?**], which replaces the function $\mathbf{Card}(x)$ with its largest convex lower bound $\|x\|_1$:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax \preceq b. \end{array} \tag{11}$$

2.3.2 Boolean least squares

The original boolean least squares problem in (2) is written:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array}$$

and we can relax it as an SDP:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(AX) + 2b^T Ax + b^T b \\ \text{subject to} & \left[\begin{array}{cc} X & x^T \\ x & 1 \end{array}\right] \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n, \end{array}$$

in the variables $x \in \mathbf{R}^n$ and $X \in \mathbf{S}_+^n$. This program then produces a lower bound on the optimal value of the original problem.

2.3.3 Partitioning and MAXCUT

The partitioning problem defined above reads:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n. \end{array} \tag{12}$$

Here, the problem is directly formulated as a QCQP and the variable x disappears from the relaxation, which becomes:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n. \end{array} \tag{13}$$

MAXCUT corresponds to a particular choice of matrix W.

2.4 Duality gap and conservatism estimate

Weak duality implies that the optimal value of the Lagrangian relaxation is a lower bound on the value of the original program. Because the dual is a convex program, under some constraint qualification conditions (see [?, §5.2.3]) there is no duality gap between the dual and the bidual, hence the duality gap between the original program and its dual gives a measure of the degree of "conservatism" in the relaxation. In some particular instances, it is possible to show that, even though the original program is not convex, the duality gap is zero and the convex relaxation produces the optimal value.

A QCQP with one constraint is a classic example (see Appendix B in [?], or [?] for other examples) and relies on the fact that the numerical range of two quadratic forms is a convex set. Hence, under some technical conditions, the programs:

$$\begin{array}{ll} \text{minimize} & x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & x^T P_1 x + q_1^T x + r_1 \leq 0, \end{array} \tag{14}$$

and

$$\begin{array}{ll} \text{maximize} & \gamma + \lambda r_1 + r_0 \\ \text{subject to} & \begin{bmatrix} (P_0 + \lambda P_1) & (q_0 + \lambda q_1)/2 \\ (q_0 + \lambda q_1)^T/2 & -\gamma \end{bmatrix} \succeq 0 \\ & \lambda \geq 0, \end{array} \tag{15}$$

in the variables $x \in \mathbf{R}^n$ and $\lambda \in \mathbf{R}$ respectively, produce the same optimal value, even if the first one is nonconvex. This result is also known as the S-procedure in control theory. The key implication here of course is that while the original program is possibly nonconvex and numerically hard, its dual is a semidefinite program, hence efficiently produces the optimum of the original problem.

3 Domain restriction & linearization

The relaxations technique detailed above produces "good" lower bounds on the optimal value but no feasible points. Here, we work on the complementary approach and try to find "good" feasible points corresponding to a local minimum. Let x_0 be an initial feasible point (this can be hard to find, see the discussion on phase I problems in §11.4 of [**?**]).

3.1 Linearization

We start by leaving all convex constraints unchanged, linearizing the nonconvex ones around the original feasible point $x^{(0)}$. Consider for example the constraint:

$$x^T P x + q^T x + r \leq 0,$$

we decompose the matrix P into its positive and negative parts:

$$P = P_+ - P_-, \quad \text{with } P_+, P_- \succeq 0.$$

The original constraint can be rewritten as

$$x^T P_+ x + q_0^T x + r_0 \leq x^T P_- x,$$

and both sides of the inequality are now convex quadratic functions. We linearize the right hand side around the point x_0

$$x^T P_+ x + q_0^T x + r_0 \leq (x^{(0)})^T P_- x^{(0)} + 2(x^{(0)})^T P_-(x - x^{(0)}).$$

The right hand side is now an affine lower bound on the original function $x^T P_- x$ (see §3.1.3 in [**?**]). This means that the resulting constraint is convex and more conservative than the original one, hence the feasible set of the new problem will be a convex subset of the original feasible set.

3.2 Iterative method

The new problem, formed by linearizing all the nonconvex constraints using the method described above, is convex and can be solved efficiently to produce a new feasible point $x^{(1)}$ with a lower objective value. If we linearize again the problem around $x^{(1)}$ and repeat the procedure, we get a sequence of feasible points with decreasing objective values.

4 Randomization and bounds on suboptimality

The Lagrangian relaxation techniques developed in §2 provided lower bounds on the optimal value of the program in (1), they did not however give any particular hint on how to compute good feasible points. Moreover, even though the relaxation's conservatism can be measured in theory by the duality gap, this gap is very hard to quantify in practice.

The semidefinite relaxation in (8) produces a positive semidefinite or covariance matrix together with the lower bound on the objective. In this section, we exploit this additional output to compute good approximate solutions with, in some cases, hard bounds on their suboptimality.

4.1 Randomization

In the last section, the original QCQP:

$$\begin{array}{ll} \text{minimize} & x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & x^T P_i x + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \end{array}$$

was relaxed into:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\ \text{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad \text{i=1,\ldots,m}, \\ & \left[\begin{array}{cc} X & x^T \\ x & 1 \end{array}\right] \succeq 0. \end{array} \tag{16}$$

The last (Schur complement) constraint:

$$\left[\begin{array}{cc} X & x^T \\ x & 1 \end{array}\right] \succeq 0,$$

being equivalent to $X - xx^T \succeq 0$. In other words, suppose x and X are the solution to the relaxed program in (16), then $X - xx^T$ is a covariance matrix.

If we pick x as a Gaussian variable with $x \sim \mathcal{N}(x, X - xx^T)$, x will solve the QCQP in (1) "on average" over this distribution, meaning:

$$\begin{array}{ll} \text{minimize} & \mathbf{E}(x^T P_0 x + q_0^T x + r_0) \\ \text{subject to} & \mathbf{E}(x^T P_i x + q_i^T x + r_i) \leq 0, \quad \text{i=1,\ldots,m}, \end{array}$$

and a "good" feasible point can then be obtained by sampling x a sufficient number of times, then simply keeping the best feasible point.

4.2 Bounds on suboptimality

In certain particular cases however, it is possible to get a hard bound on the gap between the optimal value and the relaxation result. A classic example is that of the MAXCUT bound described in [**?**] or [**?**, Th. 4.3.2]. The MAXCUT problem (5) reads:

$$\begin{array}{ll} \text{maximize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array} \tag{17}$$

its Lagrangian relaxation is then computed as in (13):

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n. \end{array} \tag{18}$$

Let X be a solution to this program, we look for a good feasible point by sampling a normal distribution $\mathcal{N}(0, X)$ with mean zero and variance given by X, and convert each sample point x to a feasible point by rounding it to the nearest value in $\{-1, 1\}$, *i.e.* taking $\hat{x} = \mathbf{sgn}(x)$. Crucially, when $\hat{x}$ is sampled using that procedure, the expected value of the objective $\mathbf{E}(\hat{x}^T W x)$ can be computed explicitly:

$$\mathbf{E}(\hat{x}^T W x) = \frac{2}{\pi} \sum_{i,j=1}^{n} W_{ij} \arcsin(X_{ij}) = \frac{2}{\pi} \mathbf{Tr}(W \arcsin(X)).$$

We are guaranteed to reach this expected value $2/\pi\, \mathbf{Tr}(W \arcsin(X))$ after sampling a few (feasible) points $\hat{x}$, hence we know that the optimal value OPT of the MAXCUT problem is between $2/\pi\, \mathbf{Tr}(W \arcsin(X))$ and $\mathbf{Tr}(WX)$.

Furthermore, with $\arcsin(X) \succeq X$ (see [**?**, p. 174]), we can simplify (and relax) the above expression to get:

$$\frac{2}{\pi} \mathbf{Tr}(WX) \leq OPT \leq \mathbf{Tr}(WX).$$

This means that the procedure detailed above guarantees that we can find a feasible point that is at most $2/\pi$ suboptimal (after taking a certain number of samples from a Gaussian distribution).

5 Numerical Examples

In this section, we work out some numerical examples.

5.1 Boolean least-squares

The problem is given by:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

with

$$\begin{aligned} \|Ax - b\|^2 &= x^T A^T A x - 2b^T A x + b^T b \\ &= \mathbf{Tr}\, A^T A X - 2b^T A x + b^T b \end{aligned}$$

where $X = xx^T$. We can express the BLS problem as

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, A^T A X - 2b^T A x + b^T b \\ \text{subject to} & X_{ii} = 1, \quad X \succeq xx^T, \quad \text{rank}(X) = 1 \end{array}$$

which is still a very hard problem...

5.1.1 Semidefinite relaxation for BLS

Using the technique in §2, we compute the Lagrangian relaxation of the BLS problem using:

$$X \succeq xx^T \iff \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0,$$

for $X \in \mathbf{S}_+^n$ and $x \in \mathbf{R}^n$, to obtain a semidefinite relaxation:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, A^T A X - 2b^T A^T x + b^T b \\ \text{subject to} & X_{ii} = 1, \quad \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \end{array} \tag{19}$$

in the variables $X \in \mathbf{S}_+^n$ and $x \in \mathbf{R}^n$. The optimal value of this SDP gives a lower bound for BLS, if the optimal matrix is rank one, we're done.

5.1.2 Interpretation via randomization

Using the procedure in §4, we sample a normal distribution $z \sim \mathcal{N}(x, X - xx^T)$, with $\mathbf{E}\, z_i^2 = 1$ and the SDP objective is $\mathbf{E}\, \|Az - b\|^2$. This suggests a simple randomized method for getting approximate solutions to the BLS problem:

- find $X^\star$, $x^\star$, optimal for the semidefinite relaxation in (19)
- generate sample points z_i from $\mathcal{N}(x^\star, X^\star - x^\star x^{\star T})$
- take $x_i = \mathbf{sgn}(z_i)$ as approximate solution of BLS (can repeat many times and take best one)

And finally, pick the best x_i as an approximate solution.

5.1.3 Example

We set up the problem as follows:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

with

- (randomly chosen) parameters $A \in \mathbf{R}^{150\times 100}$, $b \in \mathbf{R}^{150}$
- $x \in \mathbf{R}^{100}$, so feasible set has $2^{100} \approx 10^{30}$ points

and compare the various approximation techniques:

- The least-squares approximate solution: minimize $\|Ax - b\|$ s.t. $\|x\|^2 = n$, then round, yields an objective value 8.7% over SDP relaxation bound.

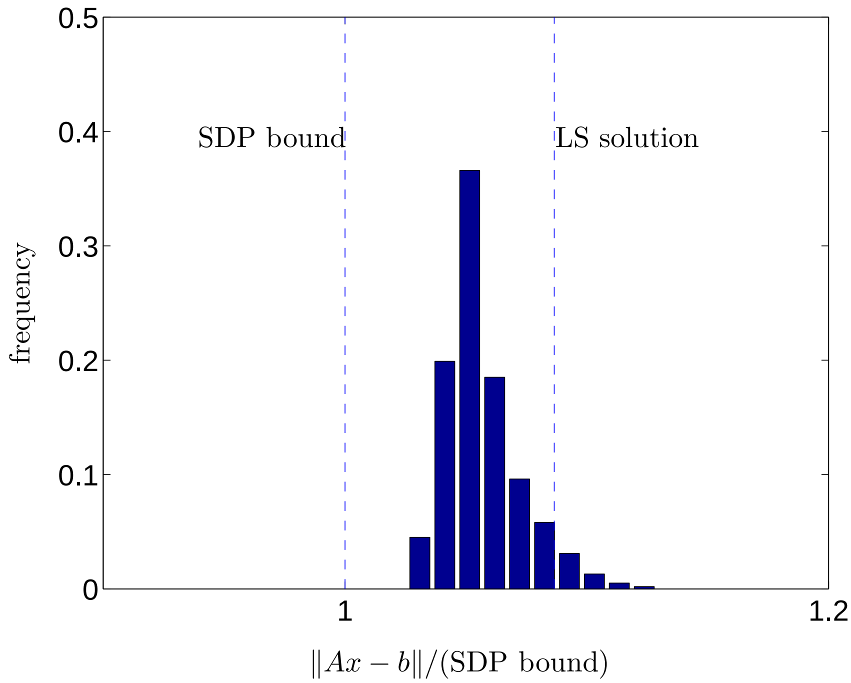

Figure 1: Distribution of objective values for points sampled using the randomization technique in §4.

- The randomized method: (using the procedure in §4)
  - best of 20 samples: 3.1% over SDP bound
  - best of 1000 samples: 2.6% over SDP bound

In figure (5.1.3), we plot the distribution of the objective values reached by the feasible points found using the randomized procedure above. Our best solution comes within 2.6% of the SDP lower bound, compared to the 8.7% of the simple LS solution.

5.2 Partitioning

We consider here the two-way partitioning problem described on pages 202–203 of [**?**], and also considered in exercise 5.39 in the same source:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array} \tag{20}$$

with variable $x \in \mathbf{R}^n$, where $W \in \mathbf{S}^n$ satisfies $W_{ii} = 0$. A feasible x corresponds to the partition

$$\{1, \ldots, n\} \;=\; \{i \mid x_i = -1\} \;\cup\; \{i \mid x_i = 1\},$$

and the matrix coefficient W_{ij} can be interpreted as the cost of having the elements i and j in the same partition, with $-W_{ij}$ the cost of having i and j in different partitions. The objective in (20) is the total cost, over all pairs of elements, and the problem (20) is to find the partition with least total cost. We define the optimal value of the partitioning problem as $p^\star$. We let $x^\star$ denote an optimal partition. (Note that $-x^\star$ is also an optimal partition.)

In §2, we have seen that the Lagrange dual of problem (20) is given by the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T\nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array} \tag{21}$$

with variables $\nu \in \mathbf{R}^n$. The dual of this SDP is the SDP

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, WX \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n, \end{array} \tag{22}$$

with variable $X \in \mathbf{S}^n$. This is the Lagrangian relaxation of problem (20). The optimal values of these two SDPs are equal, and provide a lower bound on the optimal value $p^\star$ of the partitioning problem (20). We refer to the common optimal value of the SDPs as $d^\star$ and let $\nu^\star$ and $X^\star$ denote the corresponding optimal points.

5.2.1 A simple heuristic for partitioning

One simple heuristic for finding a good partition is to solve the SDPs above, to find $X^\star$ (and the bound $d^\star$). Let v denote an eigenvector of $X^\star$ associated with its largest eigenvalue, and let $\hat{x} = \mathbf{sgn}(v)$. The vector $\hat{x}$ is our guess for a good partition.

5.2.2 A randomized method.

We generate independent samples $x^{(1)}, \ldots, x^{(k)}$ from a normal distribution on $\mathbf{R}^n$, with zero mean and covariance $X^\star$. For each sample we consider the heuristic approximate solution $\hat{x}^{(i)} = \mathbf{sgn}(x^{(i)})$. We then take the best among these, *i.e.* the one with lowest cost.

5.2.3 Greedy method

We can improve these results a little bit using the following simple greedy heuristic. Suppose the matrix $Y = \hat{x}^T W \hat{x}$ has a column j whose sum $\sum_{i=1}^n y_{ij}$ is positive. Switching $\hat{x}_j$ to $-\hat{x}_j$ will decrease the objective by $2\sum_{i=1}^n y_{ij}$. If we pick the column y_{j_0} with largest sum, switch $\hat{x}_{j_0}$ and repeat until all column sums $\sum_{i=1}^n y_{ij}$ are negative, we further decrease the objective.

5.2.4 Numerical Example

For the data in `part_prob_data.m`, the optimal SDP lower bound $d^\star$ is equal to -1641 and the $\mathbf{sgn}(x)$ heuristic gives a point (partition) with total cost -1348. Extracting a solution from the SDP solution using the simple heuristic above gives a solution with cost -1280, while applying the greedy method pushes that cost down to -1372. Exactly what the optimal value is, we can't say; all we can say at this point is that it is between -1641 and -1372.

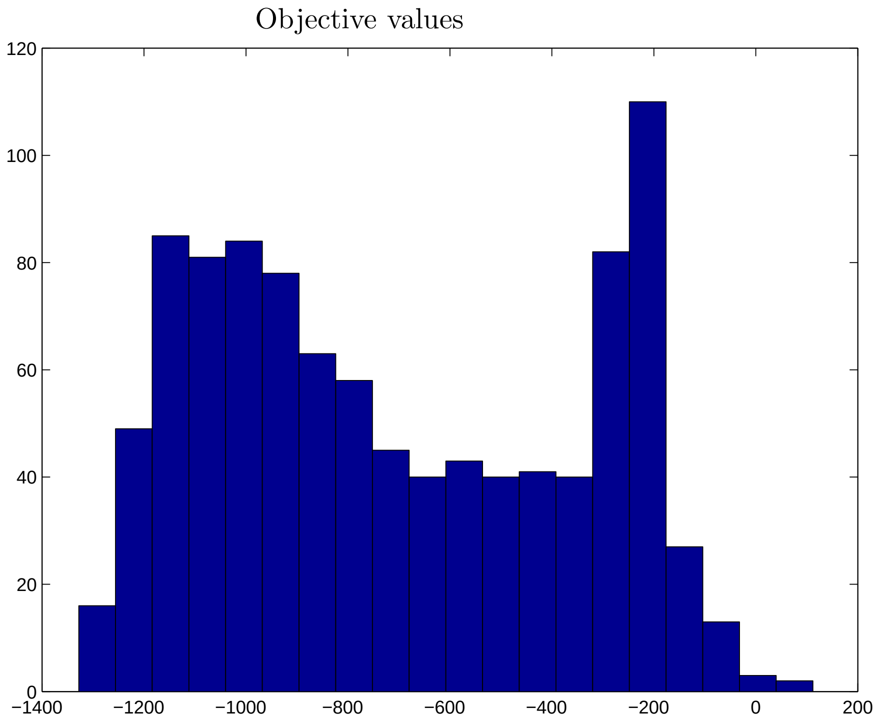

Figure 2: Histogram of the objective values attained by the random sample partitions.

We then try the randomized method, applying the greedy method to each sample, and plot in figure (5.2.4) a histogram of the objective obtained over 1000 samples. Many of these samples have an objective value larger than the original one above, but some have a lower cost. For our implementation, we found the minimum value -1392. The evolution of the minimum value found as a function of the sample size is shown in figure (5.2.4). Note that our best partition was found in around 100 samples. We're not sure what the optimal cost is, but now we know it's between -1641 and -1392.

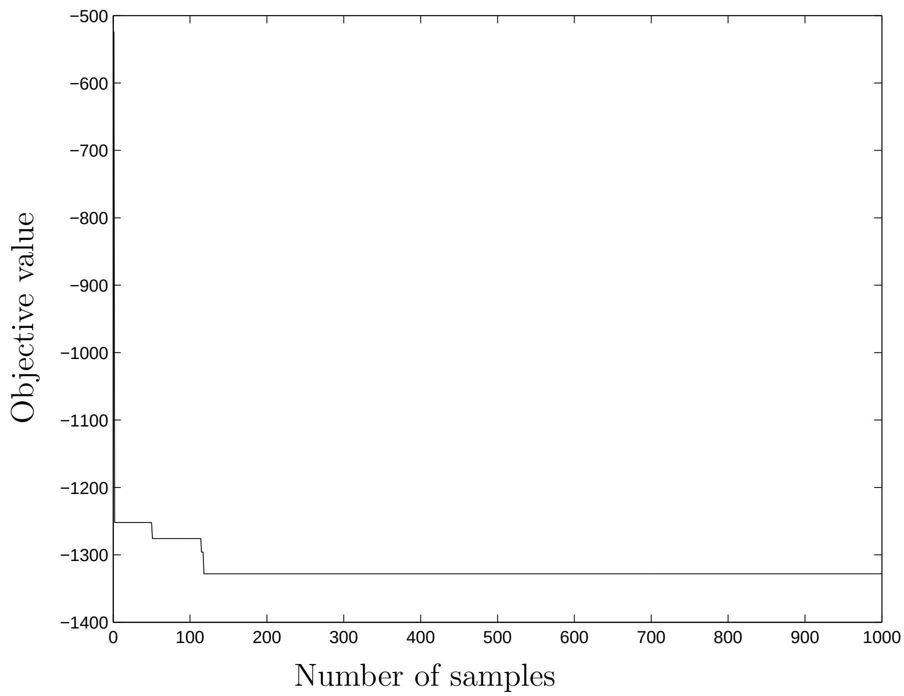

Figure 3: Best objective value versus number of sample points.