Semidefinite Relaxations and Applications

- Semidefinite relaxations
- Lagrangian relaxations for QCQPs
- Randomization
- Bounds on suboptimality
- Applications

Nonconvex problems

ee364 (more or less correct) view:

- **convex** is easy
- **nonconvex** is hard(er)

we will use convex optimization to

1. - find bounds on optimal value by **relaxation**
2. - get "good enough" feasible points by **randomization**

Basic problem: QCQPs

$$\begin{array}{ll} \text{minimize} & x^T A_0 x + b_0^T x + c_0 \\ \text{subject to} & x^T A_i x + b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m. \end{array}$$

- if all A_i are PSD, convex problem, use ee364
- here, we suppose at least one A_i is not PSD (?)

Example: Boolean Least Squares

Boolean least-squares problem is to $x_i = \pm 1,\ i=1,\dots,n$

$$\text{minimize } \|Ax-b\|_2^2 \text{ subject to } x_i^2 = 1, \quad i = 1, \ldots, n$$

- basic problem in digital communications (noisy channel)
- could check all 2^n possible values of x ...
- an NP-hard problem, and very hard in practice
- many heuristics for approximate solution $-1 \le x_i \le 1$

Example: Partitioning Problem

two-way partitioning problem (§5.1.5 in [BV04]):

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

$[x_i W_{ij} x_j]$

where $W = W^T$, $W_{ii} = 0$

- feasible $x \in \{-1, 1\}$ corresponds to partitioning

$x_i W_{ij} x_j$

- coefficients W_{ij} interpreted as the cost of having the elements i and j in the same partition.
- the objective is to find the partition with least total cost
- classic particular instance: MAXCUT $(W_{ij} \geq 0)$

Example: cardinality problems

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

introduce $z_i \in \{0, 1\}$, i.e. $z_i(1 - z_i) = 0$,

$\Leftrightarrow$

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T z \\ \text{subject to} & z_i - z_i^2 = 0, \;\; x_i(1 - z_i) = 0 \quad i = 1, \ldots, n \\ & x \in \mathcal{C} \end{array}$$

$\Rightarrow z_i = \begin{cases} 0, & x_i = 0 \\ 1, & x_i \text{ can be any value} \end{cases}$

Semidefinite relaxation

original QCQP

$$\left\{ \begin{array}{l} x^T A_0 x = \mathrm{Trace}(A_0 X) \\ X = x x^T \end{array} \right.$$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T A_0 x + b_0^T x + c_0 \\ \text{subject to} & x^T A_i x + b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m. \end{array}$$

is equivalent to

affine in X

$$\begin{array}{ll} \underset{X,\,x}{\text{minimize}} & \underline{\mathbf{Tr}(A_0 X)} + b_0^T x + c_0 \\ \text{subject to} & \mathbf{Tr}(A_i X) + b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \\ & \underline{X = x x^T} \quad \text{(non-convex)} \end{array}$$

change $X = xx^T$ into $\underline{X \succeq xx^T}$ $\Rightarrow X - xx^T \succeq 0$

relaxation

Lagrangian relaxation

original QCQP

$$\begin{array}{ll} \text{minimize} & x^T A_0 x + b_0^T x + c_0 \\ \text{subject to} & x^T A_i x + b_i^T x + c_i \le 0, \quad i = 1, \dots, m. \end{array}$$

forming Lagrangian

$$\sum_{i=1}^{m} \lambda_i c_i$$

$$L(x, \lambda) = x^T \left(A_0 + \sum_{i=1}^{m} \lambda_i A_i \right) x + \left(b_0 + \sum_{i=1}^{m} \lambda_i b_i \right)^T x + c_0 + \lambda^T c$$

recall that

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

Lagrangian relaxation: dual

$$L(x,\lambda) = x^T\Big(A_0 + \sum_{i=1}^m \lambda_i A_i\Big)x + \Big(b_0 + \sum_{i=1}^m \lambda_i b_i\Big)^T x + c_0 + \lambda^T c$$

has (for $B = [b_1 \ \cdots b_m]^T \in \mathbf{R}^{m\times n}$)

$$\begin{aligned} g(\lambda) &= \inf_x L(x,\lambda) \\ &= -\frac{1}{4}(b_0 + B^T\lambda)^T\Big(A_0 + \sum_i \lambda_i A_i\Big)^\dagger (b_0 + B^T\lambda) + \lambda^T c + c_0 \end{aligned}$$

Lagrangian relaxation: dual

Taking Schur complements gives dual problem

dual problem

$$\begin{array}{ll} \text{maximize} & \frac{1}{4}\gamma + c^T\lambda + c_0 \\ \text{subject to} & \begin{bmatrix} (A_0 + \sum_{i=1}^m \lambda_i A_i) & (b_0 + B^T\lambda) \\ (b_0 + B^T\lambda)^T & -\gamma \end{bmatrix} \succeq 0, \\ & \lambda \succeq 0 \end{array}$$

semidefinite program in variable $\lambda \in \mathbf{R}_+^m$ and can be solved "efficiently"

Lagrangian relaxation: Bidual

Taking dual again gives SDP

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(A_0 X) + b_0^T x + c_0 \\ \text{subject to} & \mathbf{Tr}(A_i X) + b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \\ & \begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix} \succeq 0 \quad \Rightarrow X - xx^T \succeq 0 \end{array}$$

in variables $X \in \mathbf{S}^n$, $x \in \mathbf{R}^n$

- have recovered original SDP relaxation "automatically"
- convexification of original problem!

(Ai at least one is not PSD)

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & x^T A_0 x \\
\text{subject to} \quad & x^T A_i x \le 0, \ i=1,\dots,m
\end{aligned}
$$

$\Downarrow$ equivalent

$$
\begin{aligned}
\underset{X}{\text{minimize}} \quad & \mathrm{Trace}(A_0 X) \\
\text{subject to} \quad & \mathrm{Trace}(A_i X) \le 0, \ i=1,\dots,m \\
& \mathrm{rank}(X) = 1, \quad X \succeq 0
\end{aligned}
$$

$\Downarrow$ drop rank-one constraint)

$$
\text{SDP} \left\{
\begin{aligned}
\text{minimize} \quad & \mathrm{Trace}(A_0 X) \\
\text{subject to} \quad & \mathrm{Trace}(A_i X) \le 0, \ i=1,\dots,m \\
& X \succeq 0
\end{aligned}
\right.
$$

original QCQP: $\textcircled{1}\begin{cases} \text{minimize} & x^T A_0 x + b_0^T x + c_0 \\ \text{subject to} & x^T A_i x + b_i^T x + c_i \le 0, \ i=1,\dots,m \end{cases}$

the Lagrange of ① is

$$L(x,\lambda) = x^T A_0 x + b_0^T x + c_0 + \sum_{i=1}^{m} \lambda_i (x^T A_i x + b_i^T x + c_i)$$

$$= x^T \underbrace{(A_0 + \sum_{i=1}^{m} \lambda_i A_i)}_{A} x + \underbrace{(b_0 + \sum_{i=1}^{m} \lambda_i b_i)}_{b}{}^T x + c_0 + \sum_{i=1}^{m} \lambda_i c_i$$

lemma 1: The optimal value of the quadratic problem

$$\boxed{\underset{x}{\text{minimize}} \quad x^T A x + b^T x + r}$$

is given by

$$p^* = \begin{cases} -\frac{1}{4} b^T A^+ b + r, & A \succeq 0, \ b \in R(A) \\ -\infty, & \text{otherwise} \end{cases}$$

where A^+ is the pseudo-inverse of A if $A \in R^{m \times n}$

$$A^+ = \begin{cases} (A^T A)^{-1} A^T, & \text{rank}(A) = n \\ A^T (A^T A)^{-1} A, & \text{rank}(A) = m \\ A^{-1}, & A \text{ is squared and non-singular} \end{cases}$$

In this case, the dual function is

$$g(\lambda) = \inf_X L(X, \lambda)$$

$$= \begin{cases} c_0 + \sum_{i=1}^m \lambda_i c_i - \frac{1}{4}\left(b_0 + \sum_{i=1}^m \lambda_i b_i\right)^T \underbrace{\left(A_0 + \sum_{i=1}^m \lambda_i A_i\right)}_{A}^{+} \underbrace{\left(b_0 + \sum_{i=1}^m \lambda_i b_i\right)}_{b}, \\ \qquad \underbrace{A_0 + \sum_{i=1}^m \lambda_i A_i}_{A} \succeq 0, \quad \underbrace{b_0 + \sum_{i=1}^m \lambda_i b_i}_{b} \in \mathcal{R}\underbrace{\left(A_0 + \sum_{i=1}^m \lambda_i A_i\right)}_{A} \\ -\infty, \text{ otherwise} \end{cases}$$

Lemma 2: Schur Complement: Consider a matrix $X \in S^n$ partitioned as

$$X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix},$$ where $A \in S^k$, if $\det(A) \neq 0$,

then the matrix $S = C - B^T A^+ B$ is called the Schur complement of A in X

The following conditions are equivalent for $X \in S^n$ ($S = C - B^T A^+ B$, A is non-singular)

(1) $X \succeq 0$

(2) $A \succeq 0$, $(I - AA^+)B = 0$, $S \succeq 0$

$$g(\lambda) = c_0 + \sum_{i=1}^{m} \lambda_i c_i - \frac{1}{4} \underbrace{\left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)^T}_{B^T} \underbrace{\left(A_0 + \sum_{i=1}^{m} \lambda_i A_i\right)^{\dagger}}_{A \succeq 0} \underbrace{\left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)}_{B}$$

$$Z = \begin{bmatrix} \underbrace{A_0 + \sum_{i=1}^{m} \lambda_i A_i}_{A} & \underbrace{b_0 + \sum_{i=1}^{m} \lambda_i b_i}_{B} \\ \left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)^T & \underbrace{-\gamma}_{C} \end{bmatrix} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

Schur complement of Z:

$$S = -\gamma - \underbrace{\left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)^T}_{B^T} \underbrace{\left(A_0 + \sum_{i=1}^{m} \lambda_i A_i\right)^{\dagger}}_{A \succeq 0} \underbrace{\left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)}_{B} \succeq 0$$

$$\Rightarrow \frac{1}{4}\gamma \le -\frac{1}{4} \underbrace{\left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)^T}_{B^T} \underbrace{\left(A_0 + \sum_{i=1}^{m} \lambda_i A_i\right)}_{A \succeq 0} \underbrace{\left(b_0 + \sum_{i=1}^{m} \lambda_i b_i\right)}_{B}$$

$$\Rightarrow g(\lambda) = \inf_{x} L(\lambda, x) = \underbrace{\frac{1}{4}\gamma + c_0 + \sum_{i=1}^{m} \lambda_i c_i}_{\text{(objective)}}$$

From lemma 2, $A \succeq 0$, $S \succeq 0 \Rightarrow \underbrace{Z \succeq 0}_{\text{(constraint)}}$

Example: Partitioning

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

no need to maintain variable x, gives relaxation (via $X = xx^T$)

matrix lifting

drop rank$(X) = 1$

$$\begin{array}{ll} \underset{X}{\text{minimize}} & \mathbf{Tr}(WX) \\ \text{subject to} & X \succeq 0, \;\; \mathbf{diag}(X) = \mathbf{1} \end{array}$$

SDP relaxation solution may not be feasible for the original problem

Feasible points?

- ~~have~~ lower bounds on optimal value of problem
- ① big question: how do we compute good feasible points?
- ② ~~can~~ we measure if our lower bound is suboptimal?

Simplest idea: randomization

original problem

$$\begin{array}{ll} \text{minimize} & x^T A_0 x + b_0^T x + c_0 \\ \text{subject to} & x^T A_i x + b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m. \end{array}$$

and relaxation

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(A_0 X) + b_0^T x + c_0 \\ \text{subject to} & \mathbf{Tr}(A_i X) + b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \\ & X - xx^T \succeq 0 \end{array}$$

- if X, x solve relaxed problem, then $\boxed{X - xx^T} \succeq 0$ can be a covariance matrix.

Gaussian randomization

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

in other words:

$$
\begin{array}{ll}
\text{minimize} & \mathbf{E}[z^T A_0 z + b_0^T z + r_0] \\
\text{subject to} & \mathbf{E}[z^T A_i z + b_i^T z + c_i] \leq 0, \quad i = 1, \ldots, m
\end{array}
$$

a good feasible point obtained by sampling enough z (often more sophisticated strategies)

Gaussian randomization

- possible to get sharper guarantees and exactly feasible points, e.g. for MAXCUT or other boolean problems
- constraint

$$x_i^2 = 1$$

  so just take $x_i = \mathbf{sign}(z_i)$

- for $\hat{x} = \mathbf{sign}(z_i)$, $z_i \sim \mathcal{N}(0, X)$, have

$$\mathbf{E}[\hat{x}_i \hat{x}_j] = \frac{2}{\pi} \arcsin(X_{ij})$$

Approximation guarantees

MAXCUT relaxation

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(WX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1}, X \succeq 0 \end{array}$$

gives

$$\mathbf{E}[\hat{x}^T W \hat{x}] = \frac{2}{\pi} \mathbf{E}[W \arcsin(X)]$$

- draw a few samples $\hat{x}$, get at least that good with high probability
- optimal value of MAXCUT is between $\frac{2}{\pi} \mathbf{Tr}(W \arcsin(X))$ and $\mathbf{Tr}(WX)$.

Better rounding (Goemans & Williamson)

suppose $W_{ij} \geq 0$, maximize

$$\sum_{ij} W_{ij}(1 - X_{ij}) \quad \text{subject to} \quad \mathbf{diag}(X) = \mathbf{1}, \; X \succeq 0$$

- sample coordinates $\hat{x}_i$ at random, get $\mathbf{Tr}(W) - \mathbf{E}[\hat{x}^T W \hat{x}] = \mathbf{Tr}(W)$, at least 50% optimal
- sample directions:

$$X_{ij} = v_i^T v_j \quad \text{with} \quad \|v_i\| = 1$$

  i.e. $X = V^T V$ by Cholesky
- draw Z uniformly at random on unit sphere, set

$$\hat{x}_i = \mathbf{sign}(Z^T v_i)$$

Better rounding (Goemans & Williamson)

expected value of cut is

$$
\begin{aligned}
\mathbf{E}[W_{ij}(1-\hat{x}_i\hat{x}_j)] &= 2W_{ij}\,\mathrm{Pr}(Z \text{ separates } v_i, v_j) \\
&= 2W_{ij}\,\mathrm{Pr}(\mathbf{sign}(v_i^T Z) \neq \mathbf{sign}(v_j^T Z)) \\
&= 2W_{ij}\frac{2\theta(v_i, v_j)}{2\pi} \\
&= \frac{2}{\pi}W_{ij}\cos^{-1}(v_i^T v_j)
\end{aligned}
$$

so

$$
\sum_{ij}\mathbf{E}[W_{ij}(1-\hat{x}_i\hat{x}_j)] = \frac{2}{\pi}\sum_{ij}W_{ij}\cos^{-1}(X_{ij})
$$

- Fact: $\cos^{-1}(t) \geq \frac{\pi}{2}\alpha(1-t)$, $\alpha \approx .87856$

Better rounding: final bound

- expected weight from random cut generated by optimal X is at least

$$\frac{2}{\pi}\sum_{ij} W_{ij}\cos^{-1}(X_{ij}) \geq \alpha \sum_{ij} W_{ij}(1 - X_{ij}) = \alpha \mathsf{SDP}^*.$$

- alternatives: if $W \succeq 0$, then (Nesterov 98)

$$\mathbf{Tr}(W \arcsin(X)) \geq \mathbf{Tr}(WX)$$

so (using earlier bound)

$$\mathsf{SDP}^* \geq \mathsf{OPT} \geq \frac{2}{\pi}\mathsf{SDP}^*$$

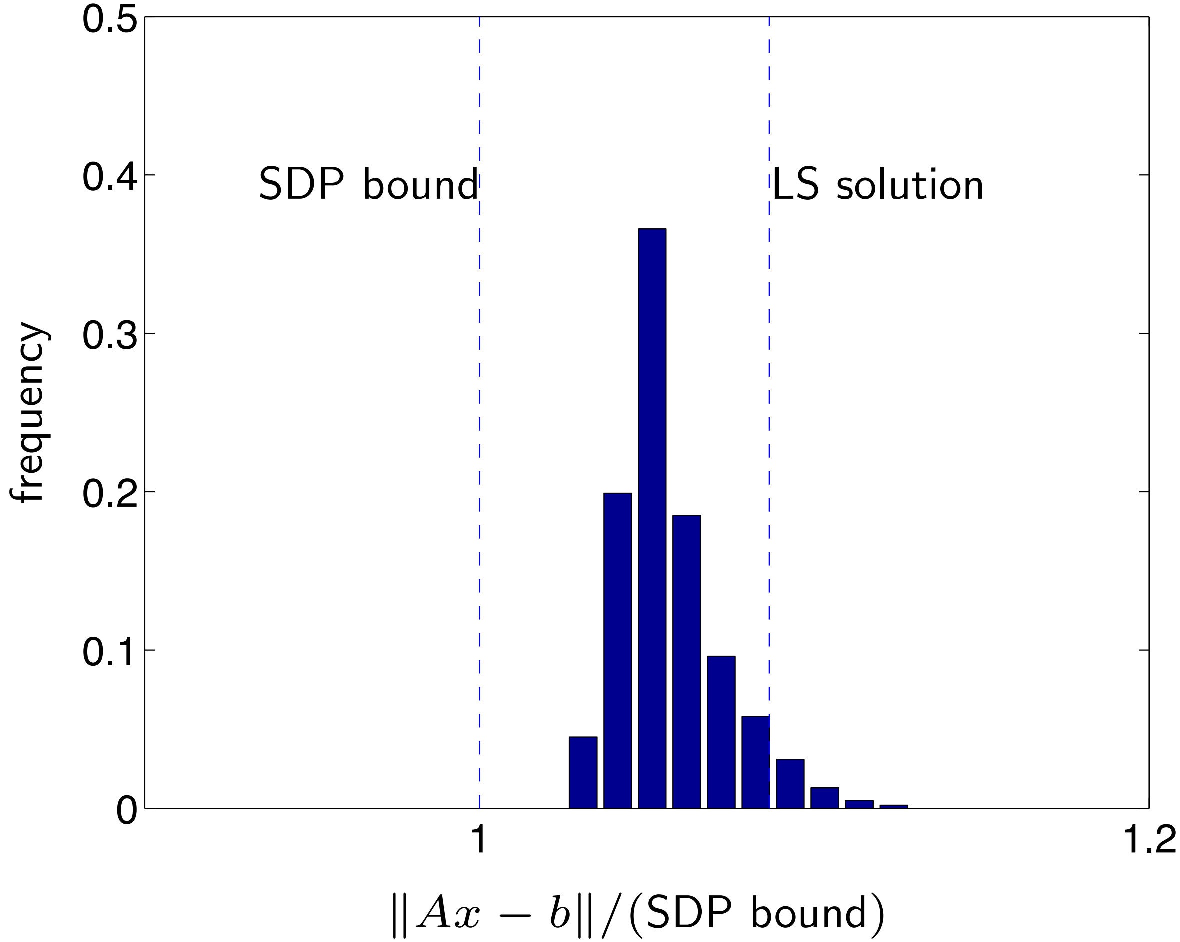
Example: boolean least squares

- (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

~~LS~~ approximate solution: minimize $\|Ax - b\|$ s.t. $\|x\|_2^2 \leq n$, then round yields objective 8.7% over SDP relaxation bound

randomized method: (using SDP optimal distribution)

- best of 20 samples: 3.1% over SDP bound
- best of 1000 samples: 2.6% over SDP bound


SDP bound
LS solution
frequency
0.5
0.4
0.3
0.2
0.1
0
1
1.2
$\|Ax - b\|/(\text{SDP bound})$

Example: partitioning problem

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

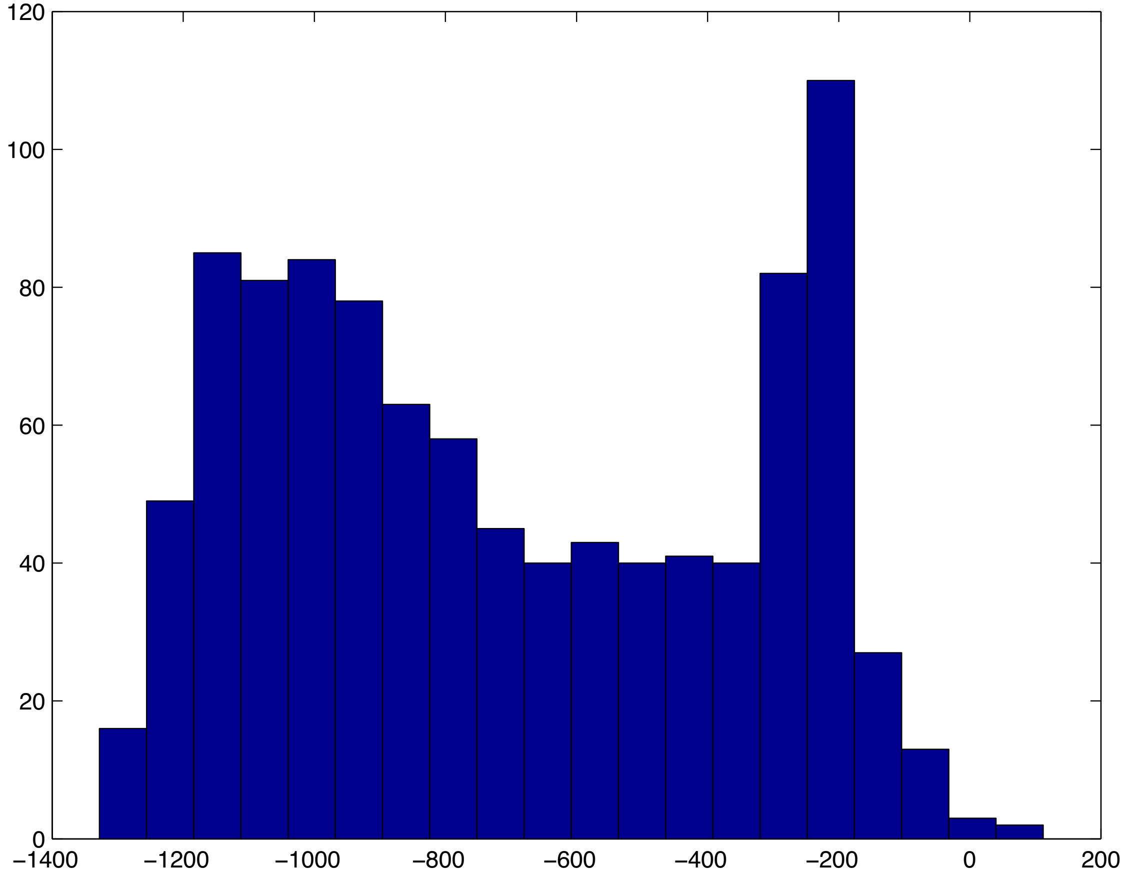
with SDP relaxation

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(WX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1}, X \succeq 0 \end{array}$$

1 and solution X^{opt} CVX

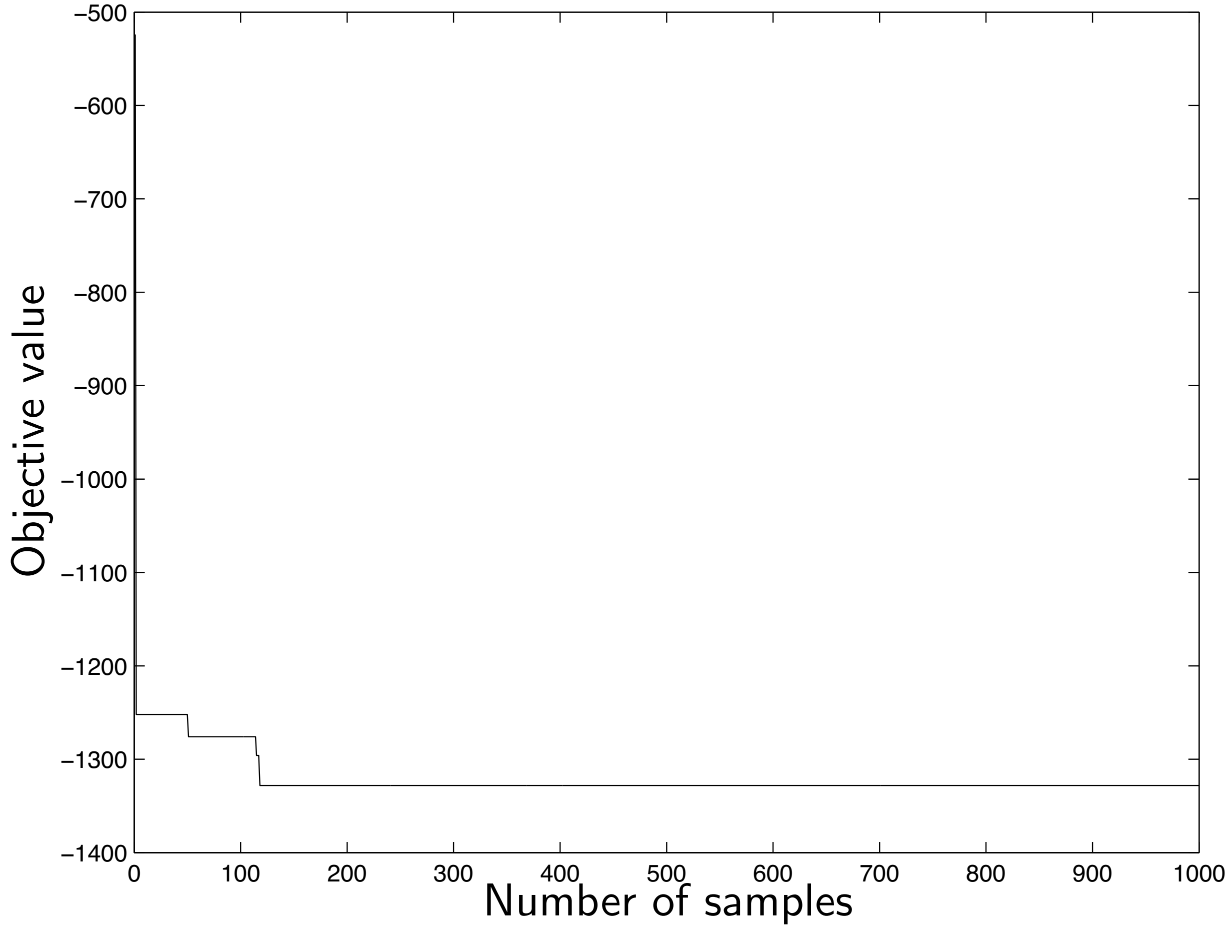
- generate samples 2 $x^{(i)} \sim \mathcal{N}(0, X^{\mathrm{opt}})$, 3 $\hat{x}^{(i)} = \mathbf{sign}(x^{(i)})$ (approximate solution)
- take one with lowest cost ($\mathrm{SDP}^{\mathrm{opt}}$ is -1641)

Histogram of partitions



heuristic on 1000 samples: minimum value attained is -1328

Objective progress in partitioning



know optimal cost is between -1641 and -1328