Convex and Nonsmooth Optimization Homework 4

Xiaoou (Charlotte) Cheng

March 1, 2022

1. (a) The Lagrangian is

$$L(x, \lambda, \nu) = c^T x + \lambda^T (Ax - b) + \sum_{i=1}^{n} \nu_i x_i (1 - x_i).$$

We derive the Lagrange dual function $g(\lambda, \nu)$. Note that for each x_i, L is at most a quadratic function. If there exists $\nu_i > 0$, then taking x_i to $+\infty$ or $-\infty$ will make $L \to -\infty$. Thus, the dual function $g = -\infty$.

Now we consider the case where $\nu_i \le 0$ for all i. If there exists an i such that $\nu_i = 0$ but $c_i + (A^T\lambda)_i \neq 0$, then we can take x_i with a different sign from $c_i + (A^T\lambda)_i$ and let its absolute value go to $+\infty$, to obtain $g = -\infty$.

The remaining case is for all i, $\nu_i \le 0$, and for all $\nu_j = 0$, the corresponding $c_j + (A^T\lambda)_j = 0$. Note that L can be perceived as "separable" in x_i, meaning that it is a sum of single variable functions of each x_i. So we can consider the infimum over each x_i separately with easy formulas. Since now for each x_i, L is convex, and the infimum is attained at

$$x_i = \frac{\nu_i + c_i + (A^T\lambda)_i}{2\nu_i}, \quad \frac{0}{0} := 0,$$

with

$$g(\lambda, \nu) = \sum_{i=1}^{n} \frac{(c_i + \nu_i + (A^T\lambda)_i)^2}{4\nu_i} - b^T\lambda$$

Therefore, the Lagrange dual function is

$$g(\lambda, \nu) = \begin{cases} -\infty, & \exists i, \nu_i > 0, \text{ or } \exists j, \nu_j = 0 \text{ and } c_j + (A^T\lambda)_j \neq 0, \\ \displaystyle\sum_{i=1}^{n} \frac{(c_i + \nu_i + (A^T\lambda)_i)^2}{4\nu_i} - b^T\lambda, & \text{otherwise.} \end{cases}$$

Therefore, the dual problem is

$$\begin{gathered} \sup\ g(\lambda, \nu), \\ \text{subject to } \lambda \ge 0. \end{gathered}$$

Actually, we can further simplify the dual problem so that we can see the connection in part (b) more clearly.

Since $\sup_{\lambda,\nu} g(\lambda, \nu) = \sup_\lambda(\sup_\nu g(\lambda, \nu))$, we can first fix λ and consider $\sup_\nu g(\lambda, \nu)$.

Note that for any i, $\frac{(c_i+\nu_i+(A^T\lambda)_i)^2}{4\nu_i} \le 0$ by the requirement $\nu_i \le 0$. Consider an arbitrary index i. If $c_i + (A^T\lambda)_i < 0$, then $\nu_i < 0$, and the function

$$\frac{(c_i + \nu_i + (A^T\lambda)_i)^2}{4\nu_i} = \frac{(c_i + (A^T\lambda)_i)^2}{4\nu_i} + \frac{\nu_i}{4} + \frac{1}{2}(c_i + (A^T\lambda)_i)$$

is maximized at

$$\nu_i = c_i + (A^T\lambda)_i,$$

with the value $c_i + (A^T\lambda)_i$.
On the other hand, if $c_i + (A^T\lambda)_i \geq 0$, then we can always take $\nu_i = -c_i - (A^T\lambda)_i$ such that

$$\frac{(c_i + \nu_i + (A^T\lambda)_i)^2}{4\nu_i} = 0.$$

Therefore, the dual problem can be simplified to be

$$\begin{aligned} \sup \quad & \sum_{i=1}^{n} \min\{c_i + (A^T\lambda)_i, 0\} - b^T\lambda \\ & \text{subject to } \lambda \geq 0. \end{aligned}$$

(b) We derive the dual of the LP relaxation.
The Lagrangian is

$$L(x, \lambda_1, \lambda_2, \lambda_3, \nu) = c^T x + \lambda_1^T(Ax - b) - \lambda_2^T x + \lambda_3^T(x - \mathbb{1}).$$

It is linear in x, so we can easily obtain the Lagrange dual function g as

$$g(\lambda_1, \lambda_2, \lambda_3, \nu) = \begin{cases} -\lambda_1^T b - \lambda_3^T \mathbb{1}, & A^T\lambda_1 = \lambda_2 - \lambda_3 - c, \\ -\infty, \text{otherwise.} \end{cases}$$

And the dual problem is

$$\begin{aligned} &\text{maximize } g(\lambda_1, \lambda_2, \lambda_3) \\ &\text{subject to } \lambda_1, \lambda_2, \lambda_3 \geq 0, \end{aligned}$$

By the definition of g, we can have that, in $\mathbf{dom} g$,

$$\lambda_3 = \lambda_2 - A^T\lambda_1 - c.$$

Thus, the simplified dual problem is

$$\begin{aligned} &\sup_{\lambda_1, \lambda_2} -\lambda_1^T b + (c + A^T\lambda_1 - \lambda_2)^T \mathbb{1}, \\ &\text{subject to } \lambda_1, \lambda_2, \lambda_2 - A^T\lambda_1 - c \geq 0. \end{aligned}$$

We want to show that it is the same problem as in the dual of the Lagrange relaxation, and the key is to express the minimum in the simplified dual problem of the Lagrange relaxation by the additional dual variable λ_2.
For given λ_1, we have $\lambda_2 \geq 0$ and $\lambda_2 \geq A^T\lambda_1 + c$. Since the objective function $-\lambda_1^T b + (c + A^T\lambda_1 - \lambda_2)^T \mathbb{1}$ is linear in λ_2, and is thus convex in λ_2, it attains maximum in λ_2 at the extreme point of the convex set $\{\lambda_2 : \lambda_2 \geq 0, \lambda_2 \geq A^T\lambda_1 + c\}$. Thus, to attain the maximum, we should take $\lambda_2 = \max\left\{A^T\lambda_1 + c, 0\right\}$, and then the objective function becomes

$$-\lambda_1^T b + (c + A^T\lambda_1 - \max\left\{A^T\lambda_1 + c, 0\right\})^T \mathbb{1} = -\lambda_1^T b + \min\left\{A^T\lambda_1 + c, 0\right\}^T \mathbb{1},$$

where taking maximum and minimum are both elementwise. Then, the dual problem becomes

$$\begin{aligned} &\sup -\lambda_1^T b + \min\left\{A^T\lambda_1 + c, 0\right\}^T \mathbb{1} \\ &\qquad \text{subject to } \lambda_1 \geq 0, \end{aligned}$$

which is exactly the dual problem we consider in (a). Therefore, their optimal values must be the same, and give the same lower bound.

2. (a) Suppose

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.$$

Then

$$\langle A_1, X \rangle = x_{11} = b_1 = 1,$$
$$\langle A_2, X \rangle = x_{22} = b_2 = 0.$$

There doesn't exist a strictly feasible $\tilde{X}$, since if $\tilde{X}$ is positive definite, then its diagonal elements must be strictly positive. However, we know that $x_{22} = 0$, so it is impossible for $\tilde{X}$ to be positive definite.

(b) $\langle C, X \rangle = x_{12} + x_{21} = 2x_{12}$.
By Sylvester's criterion for positive semidefinite matrices, a Hermitian matrix is positive semidefinite if and only if all principal minors are nonnegative. Since we know x_{11} and x_{22} already, we only need

$$\det X \geq 0,$$

which boils down to

$$-x_{12}^2 \geq 0.$$

Thus,

$$x_{12} = 0,$$

and the optimal value of the primal SDP is

$$p^* = 0,$$

which is attained at

$$X^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

(c) The Lagrangian is

$$L(X, \Lambda, \nu) = \langle C, X \rangle - \langle \Lambda, X \rangle + \sum_{i=1}^{2} \nu_i(\langle A_i, X \rangle - b_i).$$

The Lagrange dual function is

$$g(\Lambda, \nu) = \inf_{X \in S^2} L(X, \Lambda, \nu) = \begin{cases} -b^T\nu, & C - \Lambda + \sum_{i=1}^{2} \nu_i A_i = 0 \\ -\infty, & \text{otherwise.} \end{cases}$$

So the dual SDP is

$$\text{maximize } g(\Lambda, \nu),$$
$$\text{subject to } \Lambda \succeq 0,$$

which can be simplified to

$$\sup b^T y,$$
$$\text{subject to } C - \sum_{i=1}^{2} y_i A_i \succeq 0.$$

(d) We want to find y such that

$$\begin{bmatrix} -y_1 & 1 \\ 1 & -y_2 \end{bmatrix} \succ 0.$$

By Sylvester's criterion, we require

$$y_1 < 0, \quad y_1 y_2 - 1 > 0.$$

There exist a lot of $\tilde{y}$ that are strictly feasible. For example, we can take $\tilde{y} = (-1, -2)^T$.

(e) Note that $b^T y = y_1$. The constraints for the dual SDP are equivalent to (by Sylverster's criterion),

$$y_1 \leq 0, \quad y_2 \leq 0, \quad y_1 y_2 - 1 \geq 0.$$

Thus, the optimal value for the dual SDP is $d^* = 0$. It cannnot be attained, because if $y_1 = 0$ then there is no $y_2 \leq 0$ such that $y_1 y_2 - 1 \geq 0$.
The strong duality holds, because $d^* = p^* = 0$.

3. (a) If X is feasible in the matrix form, then since $\mathbf{rank} X = 1$, there exist $u, v \in \mathbb{R}^n$ such that $X = uv^T$ (we can take v to be a vector in the one-dimensional row space, and u records the coefficient needed to obtain each row in X from v^T).
Since X is symmetric, and $X_{ii} = 1$, we know that $u_1, v_1 \neq 0$. u_i/u_1, as the ratio of the i-th row to the first row, should also be v_i/v_1 by symmetry, since the latter is the raio of the i-th column to the first column. Thus, $X = uv^T = vv^T$, which is positive semidefinite.
Since $X_{ii} = v_i^2 = 1$, we have $v_i = \pm 1$. So if X is feasible, then it must have the form $X = xx^T$, where $x \in \mathbb{R}^n$ satisfies $x_i \in \{-1, 1\}$.
Therefore, for every feasible X in the matrix form, we can find $X = xx^T$ such that the objective function

$$\mathbf{tr}(WX) = \mathbf{tr}(x^T W x) = x^T W x,$$

and and the constraints $X_{ii} = 1$ become $x_i^2 = 1$. Therefore, for every feasible X in the matrix form, we can find a feasible x in the original form, and their objective functions are equivalent.
For the opposite direction, we want to prove that for every feasible x in the original form, we can find a feasible X such that their constraints and objective function follow the tranformation from x to X. If $x_i^2 = 1$, then $x_i \in \{-1, 1\}$. Let $X = xx^T$, then $x_i^2 = 1$ is equivalent to $X_{ii} = 1$. And X is a rank-1 positive semidefinite matrix. Also,

$$x^T W x = \mathbf{tr}(x^T W x) = \mathbf{tr}(W x x^T) = \mathbf{tr}(WX),$$

so their objective function also translate to each other.

(b) It gives a lower bound on the optimal value of the two-way paritioning problem, because it has less contraints than the problem in (a), while the remaining constraints and the objective function are exactly the same as in (a). The minimum over a larger feasible set is smaller, so it gives a lower bound.
If an optimal point X^* for this SDP has rank one, then it is also feasible for the original problem. which means that the minimum of the original problem must be less or equal to the optimal value of the SDP relaxtation attained at X^*. On the other hand, the optimal value of the SDP relaxation should provide a lower bound. Thus, the "equality" holds, meaning that X^* is also an optimal point for the original problem.

(c) We derive the dual problem of the SDP relaxation.
The Lagrangian is

$$L(X, \Lambda, \nu) = \langle W, X \rangle - \langle \Lambda, X \rangle + \sum_{i=1}^{n} \nu_i (X_{ii} - 1) = \langle W - \Lambda + \mathbf{diag}(\nu), X \rangle - \nu^T \mathbb{1}.$$

The Lagrange dual function is

$$g(\Lambda, \nu) = \begin{cases} -v^T \mathbb{1}, & W - \Lambda + \mathbf{diag}(\nu) = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

The dual problem can be simplified to

$$\begin{aligned} &\text{maximize} \ -\mathbb{1}^T \nu, \\ &\text{subject to } W + \mathbf{diag}(\nu) = \Lambda \succeq 0, \end{aligned}$$

which is exactly the Lagrangian relaxation.
Also, they have the nice property that the dual of the the dual is the SDP relaxation.
The Lagrangian of the Lagrange dual relaxation is

$$\tilde{L}(\nu, \Pi) = \mathbb{1}^T \nu - \langle W + \mathbf{diag}(\nu), \Pi \rangle$$

with $\Pi \in S^n$.
The Lagrange dual function of $\tilde{L}$ is

$$\tilde{g}(\Pi) = \begin{cases} -\langle W, \Pi \rangle, & \mathbf{diag}(\Pi) = \mathbb{1}, \\ -\infty, & \text{otherwise.} \end{cases}$$

Thus, the dual problem is

$$\begin{aligned} &\text{minimize} \langle W, \Pi \rangle, \\ &\text{subject to } \Pi \succeq 0, \mathbf{diag}(\Pi) = \mathbb{1}, \end{aligned}$$

which is exactly the SDP relaxation of two-way partitioning problem.
Therefore, these two SDPs are dual problems of each other.
The lower bounds found by them must be the same. By weak duality, the optimal value of the dual problem is always smaller than the optimal value of the primal problem. Since these two problems are dual problems of each other, their optimal values must be the same.

(d) **For dataset 1:** the optimal value for (5.114) is -15.0000, with optimal primal variable ν_1 being

```
    3.999998699976214e+00
    2.000003538582479e+00
   -9.999999786283382e-01
    2.137166199567275e-08
    9.999884027439561e-01
    2.137166177362815e-08
    1.000006252407510e+00
    4.000003916045245e+00
    1.999996456506412e+00
    2.000002756272565e+00
```

(the optimal dual variable Π is too large to present).
The optimal value for (5.115) is also -15.0000. The optimal dual variable ν_2 is

```
-4.000003812612343e+00
-1.999999096936917e+00
9.999999918978262e-01
-8.102173602549534e-09
-9.999871907332640e-01
-8.102173399865204e-09
-1.000006299871492e+00
-4.000002632288783e+00
-1.999997487812266e+00
-2.000003488920431e+00
```

In fact, `norm(nu2 + nu1)=7.1125e-06`, so $\nu_2 \approx -\nu_1$. Their signs change because ν_2 is the dual variable corresponding to the equality constraint, and at the optimal solution, ν_2 and $-\nu_2$ make the value of Lagrangian the same. Therefore, it is reasonable that ν_2 can change its sign from ν_1, and we can still perceive them to have correspondence.
For the dual variable Λ, from the above derivation of the dual problem of the Lagrangian relaxation, we know that it should correspond to $W + \mathbf{diag}(\nu)$ in the primal problem (5.114).
`norm(W+diag(nu1)-Lambda)= 5.1126e-06` shows this agreement.
The primal variable of (5.115), X, corresponds to the dual variable Π of (5.114). In fact, `norm(X-Pi)= ... 2.8588e-05`.

Therefore, at their optimal value, the primal variable of (5.114) corresponds to the dual variable of (5.115) with the same value. Similarly, the dual variable of (5.114) also has agreement with the primal variable of (5.115). This agreeement is expected, since for these two problems, they are the dual problem of each other, and weak duality implies strong duality.

Approximate ranks: If we treat all eigenvalues less than 10^{-7} as effectively 0, then the rank of X (and also Π) is 3, and the rank of Λ (and also $W + \mathbf{diag}(\nu)$) is 7. The sum of these two ranks is 10, which is the dimension of W. The matrices satisfy approximate comlementarity, as all entries in `(Lambda*X)` are on the order of 10^{-5}.

For dataset 2: The optimal values for both (5.114) and (5.115) are the same as -130.551, so it verifies the strong duality.

Code listings:

```
close all;
clear variables;
%%% solve (5.114)

W1=load('hw4data1.mat');
W2 = load('hw4data2.mat');

W = W2.W; % W1.W
n = size(W,1);
one = ones(n,1);

cvx_begin sdp
  variable nu(n);
  dual variable Pi
  maximize (-(one'*nu))
  - W - diag(nu) ≤ 0 : Pi;

cvx_end
```

```
close all;
clear variables;
%%% solve (5.115)

W1=load('hw4data1.mat');
W2 = load('hw4data2.mat');

W = W2.W;
n = size(W,1);
one = ones(n,1);

cvx_begin sdp
  variable X(n,n) symmetric
  dual variables Lambda nu
  minimize (trace(W*X))
  -X ≤0 : Lambda;
  diag(X)==one : nu;
cvx_end
```

(e) **Explain such an X satisfy the constraints in (5.115):**

(i) $X \succeq 0$: $X = V^T V$ is obviously symmetric. For any $y \neq 0 \in \mathbb{R}^n$,

$$y^T X y = y^T V^T V y = \|Vy\|_2^2 \geq 0.$$

(ii) $X_{ii} = 1$: $X_{ii} = v_i^T v_i = 1$.

How much the optimal value will differ: Since $y_i^2 = 1$ for any i, we have

$$\frac{1}{2}\sum_{i<j} w_{ij}(1-y_i y_j) = \frac{1}{2}\sum_{i<j} w_{ij} - \frac{1}{2}\sum_{i<j} w_{ij} y_i y_j = -\frac{1}{4} y^T W y + \frac{1}{4}\sum_{i=1}^{n} w_{ii} y_i^2 + \frac{1}{2}\sum_{i<j} w_{ij} = -\frac{1}{4} y^T W y + \frac{1}{4}\sum_{i,j} w_{ij}.$$

Thus, with the corresponding constraints, the relation between the optimal value is

$$\max\{\frac{1}{2}\sum_{i<j} w_{ij}(1-y_iy_j)\} = -\frac{1}{4}\min\{y^TWy\} + \frac{1}{4}\sum_{i,j} w_{ij}.$$

If we know the exact optimal value for either BV (5.113) or the maxcut problem, we can obtain the other optimal value from the above relation.
Similarly, the G&W SDP relaxation (P) is related to (5.115). Let $X = V^TV$.

$$\frac{1}{2}\sum_{i<j} w_{ij}(1-v_i\cdot v_j) = \frac{1}{2}\sum_{i<j} w_{ij} - \frac{1}{4}\langle W,X\rangle + \frac{1}{4}\sum_{i=1}^{n} w_{ii}|v_i|^2 = -\frac{1}{4}\langle W,X\rangle + \frac{1}{4}\sum_{i,j} w_{ij}.$$

Therefore, with all the constraints held,

$$\max\left\{\frac{1}{2}\sum_{i<j} w_{ij}(1-v_i\cdot v_j)\right\} = -\frac{1}{4}\min \mathbf{tr}(WX) + \frac{1}{4}\sum_{i,j} w_{ij}.$$

Numerical part:
For dataset 1: The assigned partition is `y =[ 1 1 1 1 -1 -1 -1 -1 1 1]`, the corresponding cut value in (Q) is $-\frac{1}{4}y^TWy + \frac{1}{4}\sum_{i,j} w_{ij} = 28$. The optimal value of the G&W SDP relaxation (P) is 29, so the corresponding cut value in (Q) is just slightly smaller than the optimal of the G& W SDP relaxation (P).
For dataset 2: The corresponding cut value in (Q) is $-\frac{1}{4}y^TWy + \frac{1}{4}\sum_{i,j} w_{ij} = 336.5$. The optimal value of the G&W SDP relaxation (P) is 348.638, so the corresponding cut value in (Q) is also just slightly smaller than the optimal of the G& W SDP relaxation (P). Code listings:

```
close all;
clear variables;

W1=load('hw4data1.mat');
W2 = load('hw4data2.mat');

W = W2.W; %W1.W
n = size(W,1);
one = ones(n,1);
one_matrix = ones(n,n);
W_sum = sum(W, 'all');


cvx_begin sdp
  variable X(n,n) symmetric
  dual variables Lambda nu
  %maximize (1/4*trace(W*(one_matrix - X)));
  maximize (-1/4*trace(W*X)+1/4*W_sum)
  X>=0 : Lambda;
  diag(X)==one : nu;
cvx_end

V = chol(X); % it doesn't have issue for W1, W2.
r = 1/sqrt(n) * ones(n,1);
S = r'* V;
% the indices with +1
idp = S>=0; % elementwise
% the indices with -1
idn = S<0;
y = ones(n,1);
y(idn) = -1;
% cut value
val = -1/4* (y'*(W*y))+ 1/4*W_sum;
```

(f) The exact value obtained is 29. We have $28/29 \approx 0.966$. It is within the range of 0.878.

The vector chosen in (e) is deterministic, though. We can generate random variables on the sphere to test the performance of this algorithm. With 100 tests, the value is 28.58, and $28.58/29 \approx 0.986$. With 1000 tests, the value is 28.65, which also lies in the range of 0.878. We also test how the empirical average of W behaves as the number of random vectors increases. The max cut value converges around 28.65, and $28.65/29 \approx 0.988$.

As an aside, to generate vectors uniformly distributed on the sphere S_n, we can use the following procedure: first let $X = (X_1, \ldots, X_n)$ be a vector of independent standard Gaussian random variables. Then, $U = X/\|X\|_2$ is uniformly distributed on S_n.

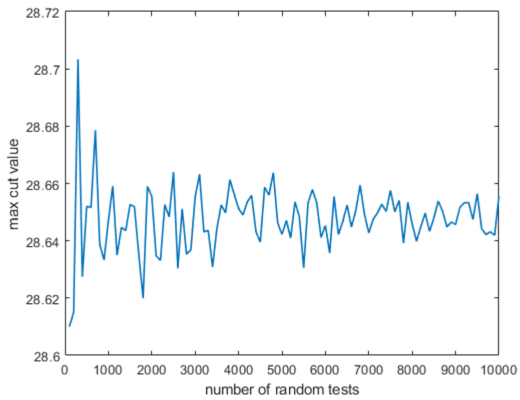

Figure 1: The empirical mean of max cut value with different numbers of random vectors

Code listings:

```
close all;
clear variables;
%%% obtain the exact value for W1

W1=load('hw4data1.mat');
W = W1.W;
n = size(W,1);

% To get all length n vectors with \pm 1,
% we map the binary representation from 0 to 2^n-1 to a vector of 0 and 1,
% then we map 0 to -1, 1 to 1 so that the vectors consist of 1 and -1.
num = 0;
y = (dec2bin(num,n) == '1');

y = y';
y = 2*y-1;



val_keep = - (y'*(W * y))/4 ; % + trace(W) / 4;
y_keep = y;

upper_bound = 2^n - 1;

while num < upper_bound
  num = num + 1;
  y = (dec2bin(num,n) == '1');

  y = y';
  y = 2*y-1;

  val_t = - (y'*(W * y))/4;

  if val_t > val_keep
```

```
        y_keep = y;
        val_keep = val_t;
    end
end

val_output = val_keep + sum(W, 'all') / 4;
```

```
close all;
clear variables;
% compare the ratio (0.878)

W1=load('hw4data1.mat');
W2 = load('hw4data2.mat');

W = W1.W;
n = size(W,1);
one = ones(n,1);
one_matrix = ones(n,n);
W_sum = sum(W, 'all');

% number of random vectors
num_list = 100:100:10000;
% record the corresponding result
val_list = zeros(size(num_list));


cvx_begin sdp
  variable X(n,n) symmetric
  dual variables Lambda nu
  maximize (-1/4*trace(W*X)+1/4*W_sum)
  %maximize (0.5*trace(W*(one_matrix - X)));
  X≥0 : Lambda;
  diag(X)==one : nu;
cvx_end

V = chol(X);

% generate r
for j = 1: length(num_list)
  vector_num = num_list(j);
  r = randn(n,vector_num);
  norm_r = vecnorm(r);
  r = r * diag((1./norm_r));
  %r = 1/sqrt(n) * ones(n,1);
  val = 0;
  for i = 1:vector_num
      S = r(:,i)'*V;
      idp = S≥0;
      idn = S<0;
      y = ones(1,n);
      y(idn) = -1;
      val = val + 1/vector_num * (-1/4*(y* (W*y')) + 1/4* W_sum);
  end
  val_list(j) = val;
end
plot(num_list, val_list, 'LineWidth',1.2);
xlabel('number of random tests');
ylabel('max cut value');
set(gcf,'Color','w');
```