

**CONVEX AND NONSMOOTH OPTIMIZATION:  
HW 4**

TERRENCE ALSUP

(1) (a) First we write down the Lagrangian

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

Note that for a minimizer to exist we need at least  $-\text{diag}(\nu) \succeq 0$ , meaning that  $\nu \leq 0$ . If  $\nu_i = 0$ , then we must also have  $(c + A^T \lambda)_i = 0$  since otherwise we would still be able to take  $x$  such that  $L(x, \lambda, \nu) \rightarrow -\infty$ . In all other cases, since this is a smooth function of  $x$ , we can take the gradient.

$$\nabla_x L(x, \lambda, \nu) = c + A^T \lambda + \nu - 2\text{diag}(\nu)x$$

Setting the gradient to zero and solving for  $x$  gives

$$x_i = \frac{(c + A^T \lambda + \nu)_i}{2\nu_i}$$

for  $i = 1, \dots, n$ . Substituting  $x$  into the Lagrangian gives the dual function

$$g(\lambda, \nu) = \begin{cases} -\lambda^T b + \sum_{i=1}^n \frac{(c + A^T \lambda + \nu)_i^2}{4\nu_i}, & \nu \leq 0, \quad (c + A^T \lambda)_i = 0 \text{ if } \nu_i = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Whenever we maximize over the dual variable  $\nu$  we can in fact maximize over each term in the sum independently. Notice however that for  $\nu_i \leq 0$

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

achieves a maximum value of 0, whenever  $c_i + (A^T \lambda)_i \geq 0$  at  $\nu_i = -(c_i + (A^T \lambda)_i)$  and a maximum value of  $c_i + (A^T \lambda)_i$  whenever  $c_i + (A^T \lambda)_i \leq 0$  at  $\nu_i = c_i + (A^T \lambda)_i$ . Written more compactly, this becomes

$$\max_{\nu_i \leq 0} \frac{(c_i + (A^T \lambda)_i + \nu_i)^2}{4\nu_i} = \min\{0, (c + A^T \lambda)_i\}$$

Thus, we can simplify the Lagrange dual problem as maximization only over the dual variable  $\lambda$ :

$$\underset{\lambda \geq 0}{\text{maximize}} \quad -\lambda^T b + \sum_{i=1}^n \min\{0, (c + A^T \lambda)_i\}$$

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

$$L(x, \lambda, \nu_0, \nu_1) = c^T x + \lambda^T (Ax - b) - \nu_0^T x + \nu_1^T (x - \mathbf{1}) = (c + A^T \lambda - \nu_0 + \nu_1)^T x - \lambda^T b - \nu_1^T \mathbf{1}$$

Since this is linear in  $x$  we can easily see that unless  $c + A^T \lambda - \nu_0 + \nu_1 = 0$  we will have  $\inf_x L(x, \lambda, \nu_0, \nu_1) = -\infty$ . Thus, the Lagrange dual function is

$$g(\lambda, \nu_0, \nu_1) = \begin{cases} -\lambda^T b - \nu_1^T \mathbf{1}, & c + A^T \lambda - \nu_0 + \nu_1 = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

The dual problem is

$$\begin{aligned} & \text{maximize} && -\lambda^T b - \nu_1^T \mathbf{1} \\ & \text{subject to} && c + A^T \lambda - \nu_0 + \nu_1 = 0 \\ & && \lambda \geq 0, \nu_0 \geq 0, \nu_1 \geq 0 \end{aligned}$$

Notice that by the equality constraint we have  $-\nu_1 = c + A^T \lambda - \nu_0$  with the requirement that  $c + A^T \lambda - \nu_0 \leq 0$ . Therefore, the problem is equivalent to

$$\begin{aligned} & \text{maximize} && -\lambda^T b + (c + A^T \lambda - \nu_0)^T \mathbf{1} \\ & \text{subject to} && \nu_0 \geq \min\{0, c + A^T \lambda\} \\ & && \lambda \geq 0 \end{aligned}$$

where the constraint on  $\nu_0$  is component-wise. Since  $\nu_0 \geq 0$  we want to take  $\nu_0$  as small as possible to maximize the objective function. In particular, we just take  $\nu_0 = \min\{0, c + A^T \lambda\}$ . Thus, the dual problem is equivalent to

$$\text{maximize}_{\lambda \geq 0} -\lambda^T b + \min\{0, c + A^T \lambda\}^T \mathbf{1} = \text{maximize}_{\lambda \geq 0} -\lambda^T b + \sum_{i=1}^n \min\{0, (c + A^T \lambda)_i\}$$

and we see that this is equivalent to the dual problem for the reformulation of the Boolean LP. Since the dual problems are equivalent, we will obtain the same lower bound on the primal Boolean LP problem.

(2) (a) Since  $X \in S^2$  we can write

$$X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix}$$

The constraint  $\langle A_1, X \rangle = b_1$  is now  $x_1 = 1$  and  $\langle A_2, X \rangle = b_2$  is now  $x_3 = 0$ . The positive semi-definite constraint just becomes  $-x_2^2 \leq 0$ . Of course, this is only possible if  $x_2 = 0$ , so in fact the entire matrix  $X$  is determined, which means the matrix

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

is the only feasible point. In this case, the Slater condition does not hold as there is no strictly feasible  $\tilde{X}$ .

- (b) The optimal value of the primal SDP is just  $p^* = 0$ , since  $\langle C, X \rangle = 2x_2 = 0$  and is attained by the matrix

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

which is the only feasible point.

- (c) To find the dual SDP we first find the Lagrangian.

$$L(X, Z, \lambda) = \langle C, X \rangle + \sum_{i=1}^2 \lambda_i (\langle A_i, X \rangle - b_i) - \langle X, Z \rangle$$

Taking the infimum over  $X \in \mathbf{S}^n$  we see that we must have

$$C - Z + \lambda_1 A_1 + \lambda_2 A_2 = 0$$

otherwise  $\inf_X L(X, Z, \lambda) = -\infty$ . Therefore, the Lagrange dual function is

$$g(Z, \lambda) = \begin{cases} -\lambda^T b, & C + \lambda_1 A_1 + \lambda_2 A_2 - Z = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

The dual SDP can now be written as

$$\begin{aligned} & \text{maximize} && -\lambda^T b \\ & \text{subject to} && C + \lambda_1 A_1 + \lambda_2 A_2 - Z = 0 \\ & && Z \succeq 0 \end{aligned}$$

Of course we can simplify this to eliminate  $Z$  since  $Z = C + \lambda_1 A_1 + \lambda_2 A_2$ . Therefore, the dual SDP is equivalent to

$$\begin{aligned} & \text{maximize} && -\lambda^T b \\ & \text{subject to} && C + \lambda_1 A_1 + \lambda_2 A_2 \succeq 0 \end{aligned}$$

However,

$$C + \lambda_1 A_1 + \lambda_2 A_2 = \begin{bmatrix} \lambda_1 & 1 \\ 1 & \lambda_2 \end{bmatrix}$$

and  $-\lambda^T b = -\lambda_1$ . The matrix above is positive semidefinite whenever  $\lambda_1 \geq 0$  and  $\lambda_1 \lambda_2 - 1 \geq 0$ . The problem could therefore be reformulated as

$$\begin{aligned} & \text{maximize} && -\lambda_1 \\ & \text{subject to} && \lambda_1 \geq 0, \quad \lambda_1 \lambda_2 \geq 1 \end{aligned}$$

- (d) In this case the Slater condition holds for the dual SDP. For example consider  $\lambda = (2, 2)$  so that

$$Z = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \succ 0$$

We can find arbitrarily many matrices that are strictly positive definite, which means that the Slater condition holds.

- (e) The optimal value of the dual SDP requires taking  $\lambda_1 \geq 0$  as small as possible, but it cannot be smaller than  $1/\lambda_2$ . Since  $\lambda_2 \geq 0$  is unbounded we can take  $\lambda_2 \rightarrow \infty$  and hence  $\lambda_1 \rightarrow 0$  so that the optimal value of the dual SDP is  $d^* = 0$ . However, this is not attained.
  - (f) Yes, in this case strong duality holds because  $p^* = d^*$  even though the optimal value of dual SDP is never actually attained.
  - (g) In general, for SDPs if the Slater condition holds for at least one of a primal-dual pair of SDPs then we will have strong duality. This is because, for SDPs, the dual of the dual problem is the primal problem and that if the Slater condition holds then the optimal value for the dual problem is attained. If the Slater condition holds for the primal problem then the dual optimum is attained. However, the dual of the dual is the primal which means the primal optimum is attained and must be equal to the dual optimum. Similarly, if the Slater condition holds for the dual problem, as it did in this example, then the primal optimum is attained and by the same reasoning we still have  $d^* = p^*$ . Either way, strong duality holds.
- (3) (a) Suppose that  $X \in \mathbf{S}^n$  is feasible for the original problem meaning that  $x_i^2 = 1$ . We have that  $X = xx^T$  is rank 1, and that  $X_{ii} = x_i x_i = x_i^2 = 1$ . Moreover,  $X \succeq 0$  because for any  $y \in \mathbb{R}^n$  we get  $y^T X y = y^T x x^T y = \|y^T x\|^2 \geq 0$ . Thus,  $X$  is a feasible point for the matrix form problem.

Similarly, suppose that  $X$  is a feasible point for the matrix form problem. Since  $\text{rank}(X) = 1$ , we know that  $X = uv^T$  for some  $u, v \in \mathbb{R}^n$  with  $u, v \neq 0$ . Since  $X$  is symmetric we know that  $uv^T = vu^T$  meaning that

$$\|u\|^2 \|v\|^2 = u^T u v^T v = u^T v u^T v = (u^T v)^2$$

In particular,  $|u^T v| = \|u\| \cdot \|v\|$  meaning we actually have equality for the Cauchy-Schwarz inequality. This is true if and only if  $v = \alpha u$  for some  $\alpha \neq 0$ . Thus we can write  $X = \alpha u u^T$ . Since  $X \succeq 0$  we must have

$$\alpha y^T u u^T y = \alpha (y^T u)^2 \geq 0$$

meaning that in fact  $\alpha > 0$ . Therefore, define  $x = \sqrt{\alpha} u$  so that  $X = x x^T$ . Since  $X_{ii} = 1$  we must have that  $x_i^2 = 1$  meaning that  $x_i \in \{-1, 1\}$  and is therefore a feasible point for the original problem.

Finally, we note that the objective functions are the same since we have

$$x^T W x = \text{tr}(x^T W x) = \text{tr}(W x x^T) = \text{tr}(W X)$$

- (b) The optimal value for the relaxed problem gives a lower bound on the optimal value of the two-way partitioning problem simply because we are minimizing

over a larger set. In other words, for the relaxed problem we do not impose any constraint on the rank of  $X$ , which is larger than the set constrained to have  $\text{rank}(X) = 1$ . If it so happens that the optimal point  $X^*$  for the SDP has rank 1, then we know that it is also the optimal point for the two-way partitioning problem since it is feasible and achieves the lowest possible bound.

- (c) To find the relation between the two lower bounds we first look at the dual problem for the SDP relaxation. The Lagrangian is

$$L(X, Z, \lambda) = \langle W, X \rangle + \sum_{i=1}^n \lambda_i (\langle A_i, X \rangle - 1) - \langle Z, X \rangle$$

where  $A_i$  is the matrix with all zeros and a 1 at the entry  $(i, i)$  so that  $\langle A_i, X \rangle = X_{ii}$ . Minimizing over  $X$  we see that the dual function is

$$g(Z, \lambda) = \begin{cases} -\lambda^T \mathbf{1}, & W + \text{diag}(\lambda) - Z = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Maximizing the dual function over  $Z \succeq 0$  and  $\lambda$  gives the equivalent problem

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

which is exactly the Lagrange dual of the two-way partitioning problem.

Similarly, we will take the dual of the Lagrange dual problem and show that it is equivalent to the SDP relaxation. The Lagrangian is now

$$L(\nu, Z) = -\mathbf{1}^T \nu - \langle W + \text{diag}(\nu), Z \rangle = -\mathbf{1}^T \nu - \langle W, Z \rangle - \sum_{i=1}^n \nu_i Z_{ii}$$

Minimizing over  $\nu$  gives the dual function

$$g(Z) = \begin{cases} -\langle W, Z \rangle, & Z_{ii} + 1 = 0 \text{ for all } i = 1, \dots, n \\ -\infty, & \text{otherwise} \end{cases}$$

Because the original problem was a maximization we must have  $Z \preceq 0$ . Take  $X = -Z$  to get that the dual problem is

$$\begin{aligned} & \text{maximize} && -\langle W, X \rangle \\ & \text{subject to} && X_{ii} = 1 \text{ for all } i = 1, \dots, n \\ & && X \succeq 0 \end{aligned}$$

Of course, maximizing the negative of something results in a minimization problem, so we really get an equivalent problem is

$$\begin{aligned} & \text{minimize} && \langle W, X \rangle \\ & \text{subject to} && X_{ii} = 1 \text{ for all } i = 1, \dots, n \\ & && X \succeq 0 \end{aligned}$$

which is exactly the SDP relaxation. Therefore, both problems are equivalent and will give the exact same lower bound on the original two-way partitioning problem.

- (a) The Matlab function `solve_sdp1.m` solves the SDP

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

For data set 1 the optimal value is computed to be  $d^* = -15$  with optimal solution

$$\nu^* = (4, 2, -1, 0, 1, 0, 1, 4, 2, 2)^T$$

For data set 2 the optimal value is computed to be  $d^* = -130.5511$ , but the optimal solution is too long to record here.

- (b) The Matlab function `solve_sdp2.m` solves the SDP relaxation to the two-way partitioning problem

$$\begin{aligned} & \text{minimize} && \text{tr}(WX) \\ & \text{subject to} && X \succeq 0 \\ & && X_{ii} = 1, \quad i = 1, \dots, n \end{aligned}$$

As expected, the optimal values for this problem are exactly the same as the optimal values found for the dual problem in the last part. Recall that we in fact showed that these problems are dual to each other so it makes sense that we would get the same values. For this problem there are two dual variables. The first is the matrix  $Z$  that corresponds to the constraint  $X \succeq 0$  and the second is a vector  $y$  corresponding to the constraints  $X_{ii} = 1$ . The computed optimal dual variables  $y^* = -x^*$ , where  $x^*$  is the optimal primal variable from the other problem. In fact, we compute  $\|y^* + x^*\| = 7.11 \times 10^{-6}$ . Similarly the optimal matrix  $X^*$  for this problem and dual variable  $Z^*$  from the other problem satisfy  $X^* = Z^*$ . Here we compute,  $\|X^* - Z^*\|_F = 2.85 \times 10^{-5}$ . We compute the approximate ranks of the computed optimal primal and dual matrices by calling `eig` in Matlab and looking at the number of approximately non-zero eigenvalues. For the optimal primal matrix  $X^*$  we find that the rank is approximately 3, and for the optimal dual matrix the approximate rank is 7. Note that these matrices are  $10 \times 10$  with  $3 + 7 = 10$  so these matrices are approximately complimentary. We compute  $\|X^* Z^*\|_F = 4.10 \times 10^{-4}$ .

- (c) If  $X = V^T V$  with  $\|v_i\|_2 = 1$ , then  $X$  will indeed satisfy the constraints for the SDP relaxed problem. We have that  $X \succeq 0$  because for any  $y$  we have

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

Moreover,  $X_{ii} = v_i \cdot v_i = \|v_i\|_2^2 = 1$ . If we solve the SDP relaxation of the two-way partition problem for  $X$ , then we next want to find a factorization  $X = V^T V$  with  $\|v_i\|_2 = 1$ . Such a  $V$  is not necessarily unique. Two natural choices would be to use the Cholesky factorization  $X = R^T R$  with  $R = V$ ,

or to use an eigenvalue decomposition to compute the symmetric square root  $M$  with  $X = M^2$  and  $M = M^T = V$ . These will indeed have columns with norm 1, since  $\|v_i\|_2 = v_i^T v_i = X_{ii} = 1$ . However,  $X$  is constrained so that  $X_{ii} = 1$ . The assignment algorithm is implemented in the Matlab function `GoemansWilliamson.m`, which returns the minimal value  $x^T W x$ , the maximal value  $\frac{1}{4} \mathbf{1}^T W \mathbf{1} - \frac{1}{4} x^T W x$  from the Goemans and Williamson paper, and the partition defined by the optimal  $x$ . Note that Goemans and Williamson consider maximizing

$$\frac{1}{2} \sum_{i < j} W_{ij} (1 - x_i x_j)$$

However, since  $W$  is symmetric, this is equivalent to

$$\frac{1}{4} \sum_{i \neq j} W_{ij} (1 - x_i x_j)$$

Since  $x_i^2 = 1$  we know that  $1 - x_i^2 = 0$  and can freely add

$$\frac{1}{4} \sum_{i \neq j} W_{ij} (1 - x_i x_j) + \sum_i W_{ii} (1 - x_i^2) = \frac{1}{4} \sum_{i,j} W_{ij} (1 - x_i x_j)$$

without having changed anything. This is exactly  $\frac{1}{4} \mathbf{1}^T W \mathbf{1} - \frac{1}{4} x^T W x$  as we claimed.

The implementation uses a Cholesky factorization to compute  $V$ . If  $X$  is exactly low rank or low-rank to machine precision, then this will not work because the Cholesky factorization requires that  $X$  is strictly positive definite (hence full-rank). To counter this the implementation adds a small multiple of the identity  $10^{-12} \times I$  to  $X$  and then computes the Cholesky factorization. For data set 1, we compute that the optimal (minimization) cut value is  $-11$  for the partition

$$x = (1, 1, 1, 1, -1, -1, -1, -1, 1, 1)^T$$

The optimal (maximization) value for the Goemans and Williamson paper is 28.

For data set 2, we compute that the optimal cut value is  $-82$  and the partition is

$$x = (1, -1, -1, 1, 1, -1, 1, -1, -1, 1, 1, -1, -1, -1, 1, -1, 1, -1, -1, 1, 1, 1, -1, 1, 1, 1, -1, 1, 1, 1)^T$$

and the corresponding Goemans and Williamson optimal value is 336.5. In both of these cases we have that  $-11 > -15$  and  $-82 > -130.551$ , which is what we would expect because the optimal value of the primal problem is lower bounded by the optimal value of the SDP relaxation. Of course, here we are not claiming that we have found the optimal partition, only that the result is reasonable.

- (d) The Matlab function `partition.m` computes the brute-force solution to the two-way partition problem by trying all possible values. In particular the implementation looks at the binary expansion for all integers from 0 to  $2^n - 1$  with zeros corresponding to  $x_i = -1$  and ones corresponding to  $x_i = 1$ . We compute that the optimal (maximum) value of  $\frac{1}{4}\mathbf{1}^T W \mathbf{1} - \frac{1}{4}x^T W x$  is 29 for data set 1. The SDP relaxation and assignment from Goemans and Williamson gave 28, which has  $28/29 \approx 0.966 > 0.878$  and is within the guaranteed approximation factor as we expected.