Convex Optimization and Modeling	Jun.-Prof. Matthias Hein
Solution of Exercise Sheet 7 - 26.05.2010	

Exercise 17 - Relaxations of integer programming problems

There are two ways to get a lower bound for the optimal value of a combinatorial optimization problem. First, the dual problem is always convex and provides a lower bound for p^* by weak duality. Second, one relaxes the constraints e.g. instead of $x \in \{0,1\}$ one allows $x \in [0,1]$ and derives a continuous optimization problem which is often convex and thus can be solved efficiently.

a. **(4 Points)** Exercise 5.13

b. **(4 Points)** Exercise 5.39

Hints:

- 5.13.a) note that the resulting Lagrangian is non-convex and thus $\nabla_x L$ is not sufficient for a global optimum.
- 5.13.b) optimize over the dual variables for the equality constraints in order to see equivalence of both dual problems.
- 5.39: Any rank-one matrix X can be written as $X = uv^T$ for some vectors u, v.

Solution:

a. • The Lagrange function of the *Lagrangian relaxation* is given as

$$L(x, \lambda, \mu) = \langle c, x\rangle + \langle \lambda, Ax - b\rangle + \sum_i \mu_i x_i(1 - x_i).$$

We have,

$$\partial_{x_r} L = c_r + (A^T\lambda)_r + \mu_r - 2\mu_r x_r.$$

Note, that $\frac{\partial^2}{\partial x_s \partial x_r} L = -2\delta_{rs}\mu_s$ and thus $\mu_s < 0$ together with $\partial_{x_r} L = 0$ is a sufficient condition for a minimum. If $\mu_r > 0$ one observes that $q(\lambda, \mu) = -\infty$ as the Lagrange function is unbounded from below. If $\mu_r = 0$, $q(\lambda, \mu) = -\infty$ if $c_r + (A^T\lambda) \neq 0$, otherwise $q(\lambda, \mu) = -\langle \lambda, b\rangle$.

Thus we have in total the dual Lagrange function

$$q(\lambda, \mu) = \begin{cases} -\langle \lambda, b\rangle + \frac{1}{4}\sum_r \frac{1}{\mu_r}(c_r + (A^T\lambda)_r + \mu_r)^2, & \text{if } \mu \preceq 0, \\ -\infty, & \text{else.} \end{cases},$$

where we use the convention that $a/0 = -\infty$ and $0/0 = 0$. In the resulting dual problem we optimize over μ and obtain,

$$\max_{\mu_r \le 0} \frac{1}{\mu_r}\big(c_r + (A^T\lambda)_r + \mu_r\big)^2 = \begin{cases} 0, & \text{if } c_r + (A^T\lambda)_r \ge 0, \\ 4(c_r + (A^T\lambda)_r), & \text{else.} \end{cases}$$

Thus we get in total the dual problem,

$$\max_\lambda \quad -\langle \lambda, b\rangle + \sum_r \min\{0, c_r + (A^T\lambda)_r\}$$

$$\lambda_r \succeq 0.$$

- The dual problem of the LP can be derived as,

$$\begin{aligned}\max_{\lambda,\mu_1,\mu_2} \quad & -\langle\lambda, b\rangle - \langle\mu_2, \mathbf{1}\rangle\\ & c + A^T\lambda - \mu_1 + \mu_2 = 0,\\ & \lambda \succeq 0,\ \mu_1 \succeq 0,\ \mu_2 \succeq 0.\end{aligned}$$

which can be further simplified by replacing μ_1,

$$\begin{aligned}\max_{\lambda,\mu_1} \quad & -\langle\lambda, b\rangle - \left\langle c + A^T\lambda - \mu_1, \mathbf{1}\right\rangle\\ & c + A^T\lambda - \mu_1 \preceq 0,\\ & \lambda \succeq 0,\ \mu_1 \succeq 0.\end{aligned}$$

which can be further reduced to,

$$\begin{aligned}\max_{\lambda} \quad & -\langle\lambda, b\rangle - \sum_r \min\{0, c_r + (A^T\lambda_r)\}\\ & \lambda \succeq 0.\end{aligned}$$

We have strong duality for LPs, so that the optimal value of the dual problem of the LP relaxation is the same as the primal LP relaxation. As the dual of the LP and the dual of the Lagrangian relaxation are the same we deduce that the obtained lower bounds are the same. Thus for this problem we observe that the lower bound obtained by solving the problem with relaxed constraints is the same as going for the convex dual problem of the hard initial optimization problem.

b.

- Any rank-one matrix X can be written as $X = uv^T$ for some vectors u, v. In order that X is positive semi-definite, for all $w \in \mathbb{R}^n$ it has to hold that,

$$\langle w, u\rangle \langle w, v\rangle \geq 0,$$

Wlog we can assume that u and v have unit-norm. Use $w = \frac{u-v}{2}$, then we obtain,

$$(1 - \langle v, u\rangle)(\langle u, v\rangle - 1) \geq 0.$$

As the first expression is non-negative and the second one non-positive, the only possibility this inequality can be fulfilled is $u = v$. Thus $X = xx^T$ for $x \in \mathbb{R}^n$ describes all rank-one matrices X which are positive semi-definite. $X_{ij} = x_i x_j$ and thus $X_{ii} = 1$ implies $x_i^2 = 1$. Moreover,

$$\langle x, Wx\rangle = \mathrm{tr}(xx^T W) = \mathrm{tr}(XW),$$

as there are no other restrictions on the vector x we thus have shown equivalence.

- As the feasible set of the relaxed problem includes the feasible set of the original problem, the optimal value has to be a lower bound for the original optimal value. If the solution of the relaxed problem has rank one, it is optimal for the original problem (if the best of the larger set lies in the smaller set, it is also the best of the smaller set).
- Note, that $\mathrm{tr}(XY) \geq 0$ if and only if $Y \succeq 0$. Thu we get the following Lagrangian,

$$L(X, \Lambda, \mu) = \mathrm{tr}\left(WX\right) - \mathrm{tr}\left(X\Lambda\right) + \mathrm{tr}\left(\mathrm{diag}(\mu)(X - \mathbb{1})\right) = \mathrm{tr}\left((W - \Lambda + \mathrm{diag}(\mu))X\right) - \mathrm{tr}\left(\mathrm{diag}(\mu)\mathbb{1}\right),$$

where we use,

$$\mathrm{tr}\left(\mathrm{diag}(\mu)(X - \mathbb{1})\right) = \sum_{i,j=1}^n \mu_i \delta_{ij}(X_{ij} - \delta_{ij}) = \sum_i \mu_i (X_{ii} - 1).$$

Thus the dual problem becomes using $\operatorname{tr}\big(\operatorname{diag}(\mu)\mathbb{1}\big) = \operatorname{tr}\big(\operatorname{diag}(\mu)\big) = \sum_{i=1}^n \mu_i = \langle \mu, \mathbf{1}\rangle$,

$$\begin{aligned} \max_{\Lambda,\mu} \quad & -\langle \mu, \mathbf{1}\rangle \\ & \Lambda \succeq 0, \\ & W - \Lambda + \operatorname{diag}(\mu) = 0. \end{aligned}$$

which can be simplified to,

$$\begin{aligned} \max_{\mu} \quad & -\langle \mu, \mathbf{1}\rangle \\ & W + \operatorname{diag}(\mu) \succeq 0, \end{aligned}$$

As Slater's condition holds we have strong duality and thus both problems are equivalent.

Exercise 18 - Differentiable approximation of l_1-norm minimization

This exercise discusses a common technique where one replaces a non-smooth objective function with a smoothed version. The critical question is how good the solution of the smoothed version is with respect to the original objective function.

a. **(4 Points)** Exercise 6.4

Hint:

- For 6.4a) you can use the following steps
  - Derive the (necessary and sufficient) condition for a minimum of

    $$\min_{x\in\mathbb{R}^n} \sum_{i=1}^m \phi\Big(\langle a_i, x\rangle - b_i\Big),$$

    where $\phi(u) = \sqrt{u^2+\varepsilon}$.
  - Derive the dual problem of

    $$\begin{aligned} \min_{x\in\mathbb{R}^n,\, y\in\mathbb{R}^m} \quad & \|y\|_1 \\ \text{subject to: } & Ax - b = y, \end{aligned}$$

    where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. You need Hölders inequality with $p = 1$ and $q = \infty$.
  - Derive from the first step a dual feasible point and use that to derive a lower bound on p^*.

Solution:

a. Let $\hat{x}$ be the solution of

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^m \phi\Big(\langle a_i, x\rangle - b_i\Big), \tag{1}$$

and introduce $\hat{r} = A\hat{x} - b$. We have to show that

$$p^* \geq \sum_{i=1}^{m} \frac{\hat{r}_i^2}{\sqrt{\hat{r}_i^2 + \varepsilon}},$$

where p^* is the optimal value of the problem $\min_x \|Ax - b\|_1$.
At the optimum $\hat{x}$ of (1) we have

$$\sum_{i=1}^{m} \frac{\langle a_i, \hat{x}\rangle - b_i}{\sqrt{(\langle a_i, \hat{x}\rangle - b_i)^2 + \varepsilon}} a_i = 0,$$

which can be rewritten using the residuals $\hat{r}$ as

$$\sum_{i=1}^{m} \frac{\hat{r}_i}{\sqrt{\hat{r}_i^2 + \varepsilon}} a_i = 0.$$

We derive the dual problem of

$$\min_{x \in \mathbb{R}^n,\, y \in \mathbb{R}^m} \|y\|_1$$
$$\text{subject to: } Ax - b = y,$$

by establishing the Lagrange function

$$L(x, y, \lambda) = \|y\|_1 + \langle \lambda, Ax - b - y\rangle .$$

Minimizing with respect to x yields the constraint $\sum_{i=1}^{m} \lambda_i a_i = 0$ and we have seen in the lecture that

$$\|y\|_1 - \sum_{i=1}^{m} \lambda_i y_i \geq \|y\|_1 - \|y\|_1 \|\lambda\|_\infty = \|y\|_1 (1 - \|\lambda\|_\infty),$$

where we have used Hölders inequality with $p = 1$ and $q = \infty$. This yields the constraint $\|\lambda\|_\infty \leq 1$. Thus in total we get the dual problem,

$$\max_{\lambda \in \mathbb{R}^m} - \sum_{i=1}^{m} \lambda_i b_i$$
$$\text{subject to } \sum_{i=1}^{m} \lambda_i a_i = 0, \qquad \text{and} \qquad |\lambda_i| \leq 1,\ i = 1, \ldots, m.$$

Now, we define $\lambda'_i = \frac{\hat{r}_i}{\sqrt{\hat{r}_i^2 + \varepsilon}}$. We know from the extremal condition for the other loss that

$$\sum_{i=1}^{m} \lambda'_i a_i = 0.$$

Moreover, it is easy to see that $|\lambda'_i| \leq 1$. Therefore λ' is dual feasible and we get

$$- \sum_{i=1}^{m} \lambda'_i b_i = \sum_{i=1}^{m} \lambda'_i \langle a_i, \hat{x}\rangle - \lambda'_i b_i = \sum_{i=1}^{m} \lambda'_i \hat{r}_i = \frac{\hat{r}_i^2}{\sqrt{\hat{r}_i^2 + \varepsilon}} \leq p^*,$$

where we have used in the second step that $\sum_{i=1}^{m} \lambda'_i a_i = 0$.

b. This follows by,

$$\frac{\hat{r}_i^2}{\sqrt{\hat{r}_i^2 + \varepsilon}} = \frac{\hat{r}_i^2}{\sqrt{\hat{r}_i^2 + \varepsilon}} - |\hat{r}_i| + |\hat{r}_i| = |\hat{r}_i| + |\hat{r}_i| \left(\frac{|\hat{r}_i|}{\sqrt{\hat{r}_i^2 + \varepsilon}} - 1 \right).$$

Noting that $\|A\hat{x} - b\|_1 = \sum_i |\hat{r}_i|$ we are done. In total we get,

$$\|A\hat{x} - b\|_1 - \sum_{i=1}^{m} |\hat{r}_i| \left(1 - \frac{|\hat{r}_i|}{\sqrt{\hat{r}_i^2 + \varepsilon}} \right) \leq p^* \leq \|A\hat{x} - b\|_1 .$$

Once we have solved the approximate differentiable problem, we can check using this inequality how far away we are from the optimal solution of the original problem.