Lagrangian Relaxation: An overview

Sources for this lecture:

- D. Bertsimas and J. Tsitsiklis: Introduction to Linear Optimization, Athena Scientific, 1997

General idea

Lagrangian relaxation is a technique well suited for problems where the constraints can be divided into two sets:

- "good" constraints, with which the problem is solvable very easily
- "bad" constraints that make it very hard to solve.

The main idea is to relax the problem by removing the "bad" constraints and putting them into the objective function, assigned with weights (the *Lagrangian multiplier*). Each weight represents a penalty which is added to a solution that does not satisfy the particular constraint.

General idea (2)

We are given the following *integer linear problem*:

$$\min c^T x \tag{4.1}$$
$$Ax \geq b \tag{4.2}$$
$$Dx \geq d \tag{4.3}$$
$$x \text{ integer} \tag{4.4}$$

with A, D, b, c, d having integer entries. Let Z_{IP} be the optimal value to the ILP above and let

$$X = \{x \text{ integral } \mid Dx \geq d\}$$

We assume that optimizing over the set X can be done very easily, whereas adding the "bad" constraints $Ax \geq b$ make the problem infeasible to solve.

General idea (3)

Therefore, we introduce a dual variable for *every* constraint of $Ax \geq b$. The vector $\lambda \geq 0$ is the vector of dual variables (the *Lagrangian multipliers*) that has the same dimension as vector b. For a fixed λ, the problem

$$\min c^T x + \lambda^T (b - Ax)$$
$$Dx \geq d$$

is introduced and its optimal value is denoted by $Z(\lambda)$. By assumption, the optimal value for the relaxed problem with a fixed vector λ can be efficiently computed and it is easy to see that $Z(\lambda)$ provides a lower bound on Z_{IP}.

General idea (4)

Lemma 1. *If* (4.1) *has an optimal solution and if* $\lambda \geq 0$*, then* $Z(\lambda) \leq Z_{IP}$.

(exercise: prove this).

Of particular interest is the tightest of all bounds, that is

$$\max Z(\lambda)$$
$$\lambda \geq 0$$

The problem above is called the *Lagrangian dual*. Let

$$Z_D = \max_{\lambda \geq 0} Z(\lambda) \tag{4.5}$$

General idea (5)

If X is a finite set, say $X = \{x^1, \ldots, x^m\}$, then $Z(\lambda)$ can also be written as

$$Z(\lambda) = \min_{i=1,\ldots,m}(c^T x^i + \lambda^T (b - Ax^i))$$

The function $Z(\lambda)$ is the minimum of a finite set of linear functions of λ and therefore it is concave and piecewise linear.

It is important to note, however, that – unlike in linear programming – integer linear programming does not have strong duality theory. This implies that the optimal value of the Lagrangian dual does not have to be the same as the optimal value of the original (primal) problem. Instead of

$$Z_D = Z_{IP}$$

the following holds:

$$Z_D \leq Z_{IP}$$

(example on the following slide and in exercises).

Figure Example Lagrange

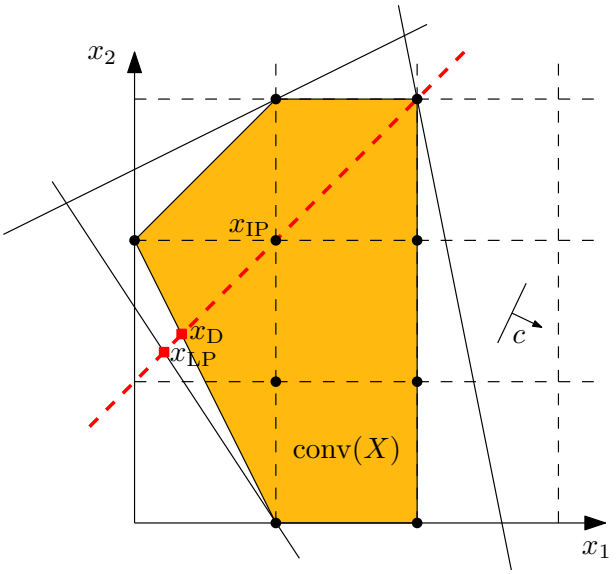
For the following example we use the following theorem:

Theorem 2. *The optimal value Z_D of the Lagrangian dual is equal to the optimal cost of the following linear programming problem:*

$$\begin{aligned} &\min c^T x \\ &\quad Ax \geq b \\ &\quad x \in \mathit{conv}(X) \end{aligned}$$

Proof omitted.

Figure Example Lagrange



Solving the Lagrangian Dual

How should the λ be set, such that the gap between $Z(\lambda)$ and Z_{IP} is as small as possible (or 0 in the best case)?

For sake of simplicity, we assume that X is finite and can be written as $X = \{x^1, \ldots, x^m\}$. Then – as described above – $Z(\lambda)$ can be written as

$$Z(\lambda) = \min_{i=1,\ldots,m}(c^T x^i + \lambda^T (b - Ax^i))$$

with $f_i = b - Ax^i$ and $h_i = c^T x^i$ this can be rewritten as

$$Z(\lambda) = \min_{i=1,\ldots,m}(h_i + \lambda f_i^T)$$

a piecewise linear and concave function.

Solving the Lagrangian Dual (2)

If $Z(\lambda)$ was differentiable then, the classical approach of maximizing the function would be the *steepest ascent method*, that is computing a sequence of iterations with

$$\lambda^{t+1} = \lambda^t + \gamma_t \nabla Z(\lambda^t).$$

We are following the gradient at the current position – with a specified stepsize γ – to reach points with a higher function value.

Unfortunately, this procedure is no longer valid for our function, since it is not differentiable everywhere.

Therefore, we extend the notion of a *gradient* in concave functions that are not differentiable everywhere. We use the following lemma:

Lemma 3. *A function $f : \Re^n \to \Re$ is concave iff for any $x^* \in \Re^n$, there exists a vector $s \in \Re^n$ such that*

$$f(x) \leq f(x^*) + s^T (x - x^*)$$

holds $\forall x \in \Re^n$.

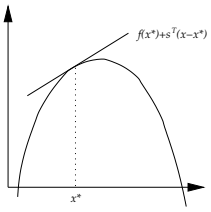
Solving the Lagrangian Dual (3)

Using the alternative notion of concavity functions, the notion of a *gradient* can be extended as follows:

Definition 4. Let f be a concave function. A vector s such that

$$f(x) \leq f(x^*) + s^T (x - x^*)$$

for all $x \in \Re^n$, is called a *subgradient* of function f at point x^*. The set of all *subgradients* of f at x^* is denoted by $\partial f(x^*)$ and is called the *subdifferential* of f at x^*.



Solving the Lagrangian Dual (4)

As soon as $0 \in \partial f(x^*)$, we are done, as the following lemma states:

Lemma 5. *Let $f : \Re^n \to \Re$ be a concave function. A vector x^* maximizes f over $\Re^n$ iff $0 \in \partial f(x^*)$.*

Lemma 5 follows directly from Lemma 3. If $(s = 0) \in \partial f(x^*)$, it implies that

$$f(x) \leq f(x^*)$$

which describes the maximum of function f.

Solving the Lagrangian Dual (5)

To get the counterpart to the *steepest ascent method* for piecewise linear functions, the notion of the *subdifferential* at a point x^* has to be extended.

Lemma 6. *Let*

$$Z(\lambda) = min_{i=1,\ldots,m}(h_i + \lambda f_i^T)$$

and

$$E(\lambda) = \{i \mid Z(\lambda) = h_i + \lambda f_i^T\}$$

Remember that $h_i = c^T x^i$ and $f_i = b - Ax^i$. Then, the following is true:

- *For every $i \in E(\lambda^*), f_i$ is a subgradient of the function Z at point λ^**
- *$\partial Z(\lambda^*) = CH(\{f_i, i \in E(\lambda^*)\})$, that is a vector s is a subgradient of the function Z at λ^* iff $Z(\lambda^*)$ is a convex combination of the vectors $f_i, i \in E(\lambda^*)$.*

Solving the Lagrangian Dual (6)

Our function is piecewise linear and concave. Hence $E(\lambda)$ is the set of indices of the linear functions that form a minimum at that specific point . In the two-dimensional case this set contains the indices of functions that cross at that specific point.

The definitions above offer the tools to generalize the *steepest ascent method* to non-differentiable, piecewise linear functions. The algorithm can be written as:

Definition 7. Subgradient optimization method.

1. Choose a starting point λ^0; $t = 0$.
2. Choose a subgradient s^t of the function Z at λ^t. If $s^t = 0 \rightarrow$ **STOP**, because the optimal value has been reached.
3. Compute $\lambda^{t+1} = \lambda^t + \gamma_t s^t$, where γ_t denotes the stepsize.
4. Increment t and go to 2.

Solving the Lagrangian Dual (7)

The definition of stepsize γ is of crucial importance, since the speed of convergence depends heavily on the stepsize. One can prove that $Z(\lambda)$ converges to the optimal value of Z – assuming it is finite – for any stepsize γ_t with

$$\sum_{t=1}^{\infty} \gamma_t = \infty$$

and

$$\lim_{t \to \infty} \gamma_t = 0 \ .$$

Such a sequence would be, *e.g.*, $\gamma_t = \frac{1}{t}$. In practice this does not always lead to quick convergence, hence other stepsize sequences are choosen.

Solving the Lagrangian Dual (8)

Held and Karp proposed the following formula for adapting the stepsize:

$$\gamma^t = \mu^t \frac{Z^* - Z(\lambda^t)}{\sum_{i=1}^{m}(b_i - \sum_{j=1}^{n} a_{ij}x^t)^2} \ ,$$

where

- Z^* is the value of the best solution for the original problem found so far
- μ^t is a decreasing adaption parameter with $0 < \mu^0 \leq 2$ and

  $$\mu^{t+1} = \begin{cases} \alpha\mu^t & Z_D \text{ did not increase in the last } T \text{ iterations} \\ \mu^t & \text{otherwise} \end{cases}$$

  with parameters $0 < \alpha < 1$ and $T > 1$.
- the denominator is the square of the length of the subgradient vector $b - Ax^t$.