Homework assignment #8

Submit answers for problem 1 only. The second problem will be discussed in the discussion sections.

1. *Proximal gradient method as Bregman proximal point algorithm* [O'Connor]. The following iteration is an extension of the proximal point algorithm (page 8.2, with $t_k = 1$) to a Bregman distance d:

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \left(f(x) + d(x, x_k)\right). \tag{1}$$

We apply this to a cost function $f(x) = g(x) + h(x)$, where g and h are convex, and g is differentiable with a Lipschitz continuous gradient. As we have seen in lecture 1 (page 1.17), this means that the function

$$\phi(x) = \frac{1}{2t} x^T x - g(x)$$

is convex for $0 < t \leq 1/L$, if L is the Lipschitz constant for the Euclidean norm.

Find the Bregman distance d generated by this kernel ϕ. Show that the proximal point iteration (1) with this distance reduces to the proximal gradient iteration

$$x_{k+1} = \operatorname{prox}_{th}(x_k - t\nabla g(x_k)).$$

2. *Exponential method of multipliers.* We consider a convex problem with m linear inequality constraints, and the dual problem:

$$\begin{array}{llll} \text{Primal:} & \text{minimize} & f(x) \\ & \text{subject to} & Ax \preceq b \end{array} \qquad \begin{array}{lll} \text{Dual:} & \text{maximize} & -b^T z - f^*(-A^T z) \\ & \text{subject to} & z \succeq 0. \end{array}$$

The dual variable z is an m-vector. In lecture 8 we interpreted the augmented Lagrangian method as the proximal point method applied to the dual problem. Here we work out what happens if we replace the squared Euclidean distance in the proximal point method with the relative entropy

$$d(u, v) = \sum_{i=1}^{m} (u_i \log(u_i/v_i) - u_i + v_i).$$

The Bregman proximal point iteration for the dual problem is

$$z_{k+1} = \underset{u}{\operatorname{argmin}} \left(b^T u + f^*(-A^T u) + \frac{1}{t_k} d(u, z_k)\right)$$

where t_k is a positive step size. Show that this is equivalent to the following iteration:

$$\begin{aligned}\hat{x} &= \operatorname*{argmin}_x \left(f(x) + \frac{1}{t_k}\sum_{i=1}^m z_{k,i}e^{t_k(a_i^T x - b_i)}\right) \\ z_{k+1,i} &= z_{k,i}e^{t_k(a_i^T \hat{x} - b_i)}, \quad i = 1, \ldots, m.\end{aligned}$$

Here a_i^T is the ith row of A, and $z_{k,i}$ is the ith component of the m-vector z_k.