Problem 1 – Convex sets

1.1 If $z_{1,2} \in A + B$ then $z_{1,2} = a_{1,2} + b_{1,2}$ with $a_{1,2} \in A$, $b_{1,2} \in B$. Hence, $tz_1 + (1-t)z_2 = [ta_1 + (1-t)a_2] + [tb_1 + (1-t)b_2]$; the first term is in $A$ and the second in $B$ by the convexity of $A, B$, therefore the sum is in $A + B$.

1.2 If $z \in S_a$, then there exist $s \in S$ so that $x = z - s$ and $||x|| \leq a$. Therefore, $S_a = S + \bar{B}(0, a)$ the (closed) ball of radius $a$ centered at the origin. Since $S$ is convex, and the ball is convex for any norm (BV), $S_a$ is convex.

1.3 The entropy is a concave function of $p$, therefore $-H(p)$ is convex, therefore the sublevel set $\{-H(p) \leq a\} = \{H(p) \geq a\}$ is convex.

1.4 The Bregman divergence is convex in $y$, with $d_\phi(y = x, x) = 0$ a minimum. Thus, the Bregman ball centered at $x$ is the sublevel set $\{d_\phi(y, x) \leq a\}$, which is convex.

Problem 2 – Boosting as Minimum Relative Entropy

\[
(MRE) \quad \min_u \sum_i u_i (\ln u_i - \ln w_i^k) \quad \text{(1)}
\]

\[\text{s.t. } \sum_i z_i u_i = 0 \quad \text{(2)}\]

\[\sum_i u_i = 1, \quad \text{(3)}\]

2.1 The objective is $\sum_i u_i (\ln u_i - \ln w_i^k) = \sum_i u_i \ln u_i - \sum_i u_i \ln w_i^k$. $u \ln u$ is known to be convex, and the second sum is a linear function in $u$, so the
objective is convex. There are only linear equality constraints, so (MRE) is a convex optimization problem.

\[ L(u, c, \nu) = \sum_i u_i (\ln u_i - \ln w_i^k) + c \sum_i z_i u_i + \nu (\sum_i u_i - 1) \]  

(4)

2.2 \[ L(u, c, \nu) = \sum_i u_i (\ln u_i - \ln w_i^k) + c \sum_i z_i u_i + \nu (\sum_i u_i - 1) \]  

(4)

2.3 \[ \frac{\partial L}{\partial u_i} = \ln u_i + 1 - \ln w_i^k + cz_i + \nu \]  

(5)

It follows that \[ u_i = w_i^k e^{-cz_i - \nu - 1} \]  

(6)

2.4 We have \[ 0 = \sum_i z_i w_i^k e^{-cz_i - \nu - 1} \]  

(7)

\[ = \sum_{z_i = +1} w_i^k e^{-\nu - 1} + \sum_{z_i = -1} (-w_i)^k e^{-\nu - 1} \]  

(8)

\[ \sum_{z_i = +1} w_i^k e^c = \sum_{z_i = -1} (-w_i)^k e^{-c - \nu - 1} \]  

(9)

\[ c = \frac{1}{2} \ln \frac{\sum_{z_i = +1} w_i^k}{\sum_{z_i = -1} w_i^k} = \frac{1}{2} \ln \frac{1 - e_k}{e_k} \]  

(10)

In the above \( e_k \) is the weighted sum of the errors of \( f_k \) and \( c \) is identical with the \( c_k \) coefficient of DISCRETEADABoost. If we plug in \( c \) in (6) and then normalize, we obtain the solution to (MRE). This solution is identical to the weight update formula for DISCRETEADABoost.

Problem 3 – General barrier function

\[ \min_x f_0(x) \]  

(11)

\[ \text{s.t. } f_i(x) \leq 0, \; i = 1 : m \]  

(12)

3.1 \( h \) is convex and increasing, and \( f_i \) is convex, which assures that \( h(f_i) \) is convex; \( f_0 \) is convex too, and \( t > 0 \). Hence, we have a linear combination of convex functions which should be convex.

3.2 Since \( x^*(t) = \min_x tf_0(x) + \phi_h(x) \), we have that the gradient of \( tf_0 + \phi_h \) vanishes at \( x^*(t) \), i.e

\[ t \nabla f_0(x^*(t)) + \sum_i h'(f_i(x^*(t))) \nabla f_i(x^*(t)) = 0 \]  

(13)

If we set now \( \lambda_i = h'(f_i(x^*(t))) / t \), this \( \lambda_i \) will satisfy \( \arg\min_x f_0 + \sum_i \lambda_i f_i(x) = x^*(t) \) hence it will be dually feasible, for primal value \( x^*(t) \).
3.3

\[ g(\lambda) = f_0(x^*(t)) + \sum_i \lambda_i f_i(x^*(t)) \]  
(14)

\[ g(\lambda) \leq p^* \leq f_0(x^*(t)) \]  
(15)

\[ \text{gap} = f_0(x^*(t)) - g(\lambda) = \frac{1}{t} \sum_i h'(f_i(x^*(t))) f_i(x^*(t)) \]  
(16)

The duality gap depends on \( u_i = f_i(x^*(t)) \). Thus we have to choose an \( h \) so that \( h'(u)u = \text{constant} \). In other words, we have to solve the differential equation

\[ \frac{dh}{du} u = C \]  
(17)

This is equivalent to \( dh = C \frac{du}{u} \) whose well known solution is \( h(u) = C \ln u + D \).

Problem 4 – Linearly Separable Support Vector Machine

Let \( g(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \). At the solution \( (w^*, \alpha^*) \), we have that \( p^* = g(\alpha^*) \) and \( w^* = \sum_i \alpha_i^* y_i x_i \). Hence,

\[ p^* = \frac{1}{2} ||w^*||^2 = \frac{1}{2} \sum_i \alpha_i \alpha_j y_i y_j x_i^T x_j \]

\[ = \sum_i \alpha_i^* - g(\alpha^*) = \sum_i \alpha_i^* - p^* \]

Therefore, \( ||w^*||^2 = 2p^* = \sum_i \alpha_i^* \).