Problem 1.

(a) First do the problem in one dimension. Let $f(x) = |x|, d(y) = \mu - \mu\sqrt{1-y^2}$.

$$\begin{aligned} f^*(y) &= \delta_{[-1,1]}(y) \\ g(y) &= f^*(y) + \mu - \mu\sqrt{1-y^2} \\ g^*(x) &= \sup_{-1\le y\le 1} xy - \mu + \mu\sqrt{1-y^2}. \end{aligned}$$

Differentiate the last expression with regard to y, we get $x - \frac{\mu y}{\sqrt{1-y^2}} = 0$. From here we deduce the optimizer y^* satisfies that $sgn(x) = sgn(y^*)$. Solve the quadratic equation

$$(1-y^2)x^2 = \mu^2 y^2,$$

one get $y^* = \frac{x}{\sqrt{x^2+\mu^2}}$; check that $y^* \in [-1, 1]$, indeed. The optimal value is

$$xy^* - \mu + \mu\sqrt{1-y^{*2}} = \frac{x^2}{\sqrt{x^2+\mu^2}} - \mu + \mu\sqrt{\frac{\mu^2}{x^2+\mu^2}} = \sqrt{x^2+\mu^2} - \mu.$$

Taking the result above, we can now conclude for n-dimensional case, $f(x) = \|x\|_1, d(y) = \mu\sum_{i=1}^{n}(1 - \sqrt{1-y_i^2})$, and

$$(f^* + d)^*(x) = \sum_{i=1}^{n}(\sqrt{x_i^2+\mu^2} - \mu).$$

(b) First observe that

$$f(x) = \max_{i=1,\cdots,n} x_i = \max_{\mathbf{1}^T p=1, p\succeq 0} p^T x.$$

We know the Legendre transform of a supporting function is the indicator functions. Moving forward,

$$\begin{aligned} g(y) &= \delta_{\mathbf{1}^T y=1, y\succeq 0}(y) + \mu\left(\sum_{i=1}^{n} y_i \log y_i + \log n\right) \\ g^*(x) &= \sup_{\mathbf{1}^T y=1, y\succeq 0} x^T y - \mu\sum_{i=1}^{n} y_i \log y_i - \mu\log n \end{aligned}$$

Write down the Lagrangian for this constrained optimization,

$$\mathcal{L}(y, \lambda, \nu) = x^T y - \mu\sum_{i=1}^{n} y_i \log y_i - \mu\log n - \lambda^T y + \nu(\mathbf{1}^T y - 1).$$

The KKT conditions are

$$\frac{\partial \mathcal{L}}{\partial y} = x - \mu(\log y - \mathbf{1}) - \lambda + \nu \mathbf{1} = 0$$
$$\lambda \succeq 0, \quad y \succeq 0$$
$$\mathbf{1}^T y = 1, \quad \lambda^T y = 0$$

We have $y = \exp\left(\frac{1}{\mu}\left(x - \mu\mathbf{1} - \lambda + \nu\mathbf{1}\right)\right)$. Note that $y \succ 0$ and hence $\lambda = 0$. Moreover,

$$\begin{aligned}
1 &= \mathbf{1}^T y \\
&= \sum_{i=1}^n \exp\left(\frac{1}{\mu}(x_i - \mu + \nu)\right) \\
&= \exp\left(\frac{\nu}{\mu}\right) \sum_{i=1}^n \exp\left(\frac{x_i - \mu}{\mu}\right) \\
\nu &= -\mu \log\left(\sum_{i=1}^n \exp\left(\frac{x_i - \mu}{\mu}\right)\right) \\
y_i &= \exp\left(\frac{x_i - \mu + \nu}{\mu}\right) \\
g^*(x) &= \sum_{i=1}^n x_i y_i - \mu \sum_{i=1}^n y_i \log y_i - \mu \log n \\
&= \sum_{i=1}^n y_i (x_i - \mu \log y_i) - \mu \log n \\
&= \sum_{i=1}^n y_i \left(\mu - \nu\right) - \mu \log n
\end{aligned}$$

Problem 2. ***Projection on order cone.***

(a) The Lagrangian for the constraint optimization problem (2) is

$$\mathcal{L}(x, z) = \frac{1}{2}\|x - a\|_2^2 + z^T A x.$$

The KKT conditions are

$$\frac{\partial \mathcal{L}}{\partial y} = x - a + A^T z = 0$$
$$z \succeq 0, \quad Ax \preceq 0$$
$$z^T A x = 0$$

(b) Observe that the algorithm sweeps through $1, \cdots, n$ and forms $\{i\}$ for each $i = 1, \cdots, n$ and merges them with successive sets based on a merging condition. This ensures the partitioning of $\{1, \cdots, n\}$ by $\beta_1, \cdots, \beta_l$. Due on the merging condition, we always have $\text{avg}(a_{\beta_k}) < \text{avg}(a_{\beta_{k+1}})$ because otherwise β_{k+1} would have been merged into β_k. To show that $\text{cs}(a_{\beta_k}) \succeq 0$, we need to show $cs(a_{\beta_k})_j \geq 0$ for $j \leq |\beta_k|$ except the last entry. Denote the subset of β_k containing it's first j indices by $\beta_k(j)$. Fix $j < |\beta_k|$, use the merging condition for the inequalities in the following derivation:

$$\begin{aligned}
\text{cs}(a_{\beta_k})_j &= \sum_{i \in \beta_k(j)} a_i - j\text{avg}(a_{\beta_k}) \\
&= j(\text{avg}(a_{\beta_k(j)}) - \text{avg}(a_{\beta_k})) \\
&\geq j(\text{avg}(a_{\beta_k(j+1)}) - \text{avg}(a_{\beta_k})) \\
&\geq \cdots \geq j(\text{avg}(a_{\beta_k(|\beta_k|)} - \text{avg}(a_{\beta_k}) = 0.
\end{aligned}$$

(c) We showed $Ax \preceq 0$, or equivalently $x_1 \leq \cdots \leq x_n$ in ii, $z \succeq 0$ in iii. Observe that for $i = 1, \cdots, n-1$,

$$(Ax)_i = \begin{cases} \text{avg}(a_{\beta_k}) - \text{avg}(a_{\beta_{k+1}}), & i = \max \beta_k \\ 0, & \text{otherwise} \end{cases}, \qquad z_i = \begin{cases} 0, & i = \max \beta_k \\ \text{(some value)}, & \text{otherwise} \end{cases}.$$

From here we deduce that $z^T Ax = 0$ indeed. Moreover, observe that for $i = 1, \cdots, n-1$,

$$(A^T z)_i = z_i - z_{i+1} = a_i - \text{avg}(a_{\beta_k}), i \in \beta_k;$$

therefore, $x + A^T z = a$.

(d) The algorithm can be implemented at $\mathcal{O}(n)$ by keeping track of the average of each β_k: in each of the n iterations described at the top of page 2, one compare $\text{avg}(a_{\beta_{l-1}})$ with a_i, if the merging condition is satisfied, keep track of the index and update the average by

$$\text{avg}(a_{\beta_{l-1}}) = \frac{|\beta_{l-1}|\text{avg}(a_{\beta_{l-1}}) + a_i}{|\beta_{l-1}| + 1}.$$