Proof for Lemma 1:

The following gave out the overall structure of Lemma 1 result/statement.

Given the mixture policy $\pi_t^{\mu}(y)$, the KL divergence between π and π_t^{μ} is given by:

$$KL(\pi, \pi_t^{\mu}) = \sum_y \pi(y) \log\left(\frac{\pi(y)}{\pi_t^{\mu}(y)}\right)$$

By substituting the definition of $\pi_t^{\mu}(y)$ into the KL divergence:

$$KL(\pi, \pi_t^{\mu}) = \sum_y \pi(y) \log\left(\frac{\pi(y)Z}{(\pi_t(y))^{1-\eta\tau}(\mu(y))^{\eta\tau}}\right)$$

Breaking down the log gives us:

$$\begin{aligned} KL(\pi, \pi_t^{\mu}) &= \sum_y \pi(y)\log(\pi(y)) - \sum_y \pi(y)(1-\eta\tau)\log(\pi_t(y)) - \sum_y \pi(y)\eta\tau\log(\mu(y)) + \sum_y \pi(y)\log(Z) \\ &= -\sum_y \pi(y)(1-\eta\tau)\log(\pi_t(y)) - \sum_y \pi(y)\eta\tau\log(\mu(y)) + \sum_y \pi(y)\log(Z) \end{aligned}$$

The term $\sum_y \pi(y)\log(\pi(y))$ is not equal to 1. However, the term $\sum_y \pi(y)\log(Z)$ simplifies to $\log(Z)$ since $\sum_y \pi(y) = 1$.

Now, to correct this, let's address the terms involving the summations over policies π, π_t, and μ:

$$\begin{aligned} -\sum_y \pi(y)(1-\eta\tau)\log(\pi_t(y)) &= (1-\eta\tau)(-\sum_y \pi(y)\log(\pi_t(y))) = (1-\eta\tau)KL(\pi, \pi_t) \\ -\sum_y \pi(y)\eta\tau\log(\mu(y)) &= \eta\tau(-\sum_y \pi(y)\log(\mu(y))) = \eta\tau KL(\pi, \mu) \end{aligned}$$

Finally, substituting these into our KL divergence gives us:

$$KL(\pi, \pi_t^{\mu}) = (1-\eta\tau)KL(\pi, \pi_t) + \eta\tau KL(\pi, \mu) + \log(Z)$$

This expression now correctly accounts for the individual terms and the normalization constant Z.

Some observation:

Z is the normalization term which is always less than or equal to 1, so the term $\log(\mathrm{Z})$ will always be non-positive

The following explains how to obtain the first term and second term of the last maths expression in the previous image.

$KL(p\|q) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right)$

When we are considering the KL divergence $KL(\pi, \pi_t)$, it is calculated as:

$KL(\pi, \pi_t) = \sum_y \pi(y) \log\left(\frac{\pi(y)}{\pi_t(y)}\right)$

Breaking down this expression, we get:

$KL(\pi, \pi_t) = \sum_y \pi(y) \log(\pi(y)) - \sum_y \pi(y) \log(\pi_t(y))$

Now, this KL divergence is part of the original expression that we are trying to simplify:

$KL(\pi, \pi_t^\mu) = \sum_y \pi(y) \log(\pi(y)) - \sum_y \pi(y)(1-\eta\tau) \log(\pi_t(y)) - \sum_y \pi(y)\eta\tau \log(\mu(y)) + \sum_y \pi(y) \log(Z)$

Notice that the first term $\sum_y \pi(y) \log(\pi(y))$ is the negative entropy of the distribution π, which is represented by $-H(\pi)$. When we subtract the second term from the first term in the expression for $KL(\pi, \pi_t)$, the term $\sum_y \pi(y) \log(\pi(y))$ (which is $-H(\pi)$) cancels out, leaving us with the KL divergence itself.

So, $(1-\eta\tau)KL(\pi, \pi_t)$ would actually expand to:

$(1-\eta\tau)KL(\pi, \pi_t) = (1-\eta\tau)\left(\sum_y \pi(y) \log(\pi(y)) - \sum_y \pi(y) \log(\pi_t(y))\right)$

$= (1-\eta\tau)\sum_y \pi(y) \log(\pi(y)) - (1-\eta\tau)\sum_y \pi(y) \log(\pi_t(y))$

This $\sum_y \pi(y) \log(\pi(y))$ term is exactly the same as the $\sum_y \pi(y) \log(\pi(y))$ term that appeared in our original $KL(\pi, \pi_t^\mu)$ expression, but multiplied by $(1-\eta\tau)$. So, when we plug this back into our $KL(\pi, \pi_t^\mu)$ expression, the $(1-\eta\tau)\sum_y \pi(y) \log(\pi(y))$ part of $(1-\eta\tau)KL(\pi, \pi_t)$ will cancel out with the $\sum_y \pi(y) \log(\pi(y))$ term from $KL(\pi, \pi_t^\mu)$.

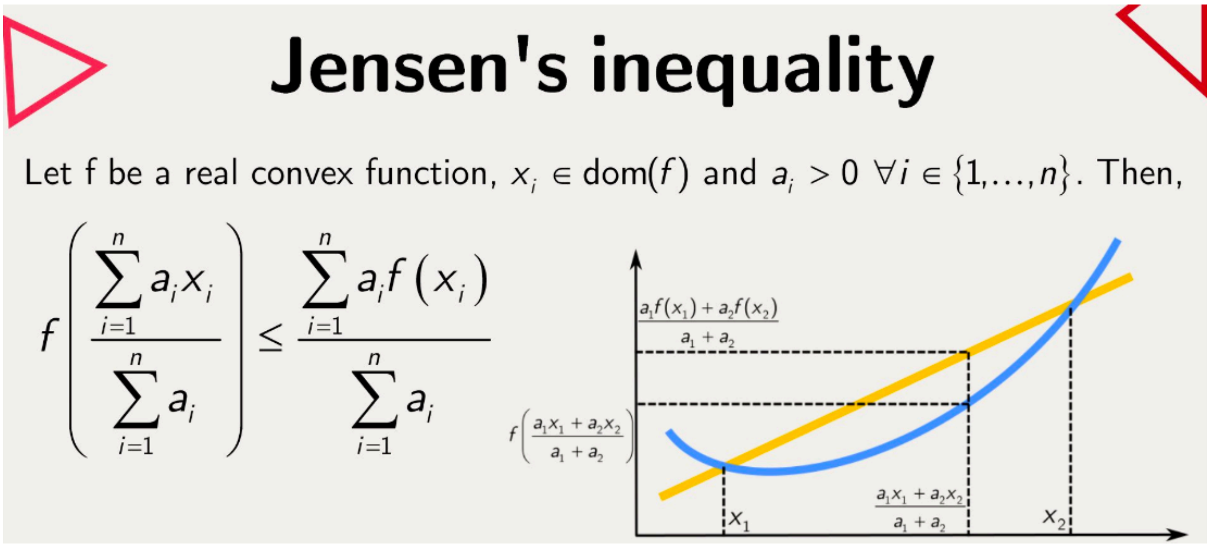
In summary, the term $\sum_y \pi(y) \log(\pi(y))$ does not appear in the final form of the lemma because it is canceled out when subtracted part of the KL divergence terms when these terms are broken down into their components.

The above only works out the maths for the first term of Lemma 1 result. As we similarly work out the maths for the second term of Lemma 1 result using the same method as above, we could actually see that the summation term for $\Sigma\, \pi(y) \log(\pi(y))$ will be cancelled out during the process of maths expansion for the coefficients of both the first term and second term of Lemma 1 result.

Here we look into the use of Jensen's inequality within Lemma 1.

$$
\begin{aligned}
\eta\tau \mathrm{KL}(\pi_t^{\mu}, \mu) &= \eta\tau \sum_y \pi_t^{\mu}(y) \log \frac{(\pi_t(y))^{1-\eta\tau} (\mu(y))^{\eta\tau}}{Z\mu(y)} \\
&= (1-\eta\tau) \sum_y \pi_t^{\mu}(y) \log \frac{(\pi_t(y))^{\eta\tau}}{(\mu(y))^{\eta\tau}} - \eta\tau \log Z \\
&\leq (1-\eta\tau) \log \sum_y \pi_t^{\mu}(y) \frac{(\pi_t(y))^{\eta\tau}}{(\mu(y))^{\eta\tau}} - \eta\tau \log Z \\
&= (1-\eta\tau) \log \sum_y \frac{(\pi_t(y))^{1-\eta\tau} (\mu(y))^{\eta\tau}}{Z} \frac{(\pi_t(y))^{\eta\tau}}{(\mu(y))^{\eta\tau}} - \eta\tau \log Z \\
&= -\log Z,
\end{aligned}
$$

Note: $\log()$ is concave, not convex. So for applying Jensen's inequality within Lemma 1, flip the sign of the inequality in the following screenshot:



As shown in the following GPT4 reply, we can also understand a bit of the maths before Jensen's inequality is being applied in Lemma 1. The maths after applying Jensen's inequality is trivial to read and understand with some sense of linear algebra manipulation.

ChatGPT

Certainly, let's walk through the mathematical steps to go from the first line to the second line in the provided equation related to the KL divergence term $\eta\tau KL(\pi_t^\mu, \mu)$.

The first line of the equation is:

$$\eta\tau KL(\pi_t^\mu, \mu) = \eta\tau \sum_y \pi_t^\mu(y) \log\left(\frac{\pi_t^\mu(y)}{\mu(y)}\right)$$

This represents the weighted KL divergence between the policy π_t^μ and the policy μ, where π_t^μ is a mixture policy, and $\eta\tau$ is the weight applied to the KL divergence.

The second line expands the definition of π_t^μ using its formula:

$$\pi_t^\mu(y) = \frac{(\pi_t(y))^{1-\eta\tau}(\mu(y))^{\eta\tau}}{Z}$$

Here, Z is a normalization constant to ensure that π_t^μ sums to 1 over all possible outputs y.

Substituting $\pi_t^\mu(y)$ into the first line gives us the second line:

$$\eta\tau \sum_y \pi_t^\mu(y) \log\left(\frac{\pi_t^\mu(y)}{\mu(y)}\right)$$

Next, we break the logarithm into parts using logarithm properties:

$$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$$

So the equation becomes:

$$= \eta\tau \sum_y \left(\frac{(\pi_t(y))^{1-\eta\tau}(\mu(y))^{\eta\tau}}{Z}\right) \left(\log\left((\pi_t(y))^{1-\eta\tau}(\mu(y))^{\eta\tau}\right) - \log(Z) - \log(\mu(y))\right)$$

Now, distribute the $\log$ function over the multiplication inside the log (using $\log(ab) = \log(a) + \log(b)$):

$$= \eta\tau \sum_y \left(\frac{(\pi_t(y))^{1-\eta\tau}(\mu(y))^{\eta\tau}}{Z}\right) \left((1-\eta\tau)\log(\pi_t(y)) + \eta\tau\log(\mu(y)) - \log(Z) - \log(\mu(y))\right)$$

Here we notice that $\eta\tau\log(\mu(y)) - \log(\mu(y))$ can be combined:

$$= \eta\tau \sum_y \left(\frac{(\pi_t(y))^{1-\eta\tau}(\mu(y))^{\eta\tau}}{Z}\right) \left((1-\eta\tau)\log(\pi_t(y)) + (\eta\tau - 1)\log(\mu(y)) - \log(Z)\right)$$

This is the simplification from the first line to the second line as seen in the screenshot. The term involving $\log(Z)$ remains constant for all y and is factored out of the summation as a constant multiple of the sum of probabilities $\pi_t^\mu(y)$, which sums to 1, giving $-\eta\tau\log(Z)$.