1. Suppose a discriminator becomes increasingly confident in distinguishing real data from generated samples. Analyze how this affects the gradient received by the generator. What does this imply about the training dynamics in early GAN iterations? Mathematically derive the generator's gradient when the discriminator is near optimal.

2. You're given two loss formulations for the generator: $\mathrm{L}_1 = E_{z\sim p(z)}[\log(1-D(G(z)))]$
$\mathcal{L}_2 = -E_{z\sim p(z)}[\log D(G(z))]$ Explain how switching from $\mathcal{L}_1$ to $\mathcal{L}_2$ changes the gradient behavior. What divergence does each implicitly minimize?

3. In training VAEs, increasing the weight of the KL-divergence term often leads to blurry reconstructions. Provide a mathematical explanation for this behavior based on the ELBO formulation. Derive the limiting behavior of the decoder's output as KL becomes dominant.

4. Imagine trying to optimize the ELBO directly with respect to the encoder parameters ϕ using naive Monte Carlo sampling. Explain why this leads to high-variance gradients. Show how a suitable change of variables enables low-variance gradients, and derive the new expression.

5. In VAEs, both the approximate posterior $q(z|x) = \mathcal{N}(\mu, \sigma^2)$ and the prior $p(z) = \mathcal{N}(0, I)$ are Gaussian. Derive the closed-form expression for the KL divergence between these distributions. Why is this derivation beneficial during training?

6. Mode collapse is a common issue in GAN training. Use a second-order Taylor approximation or Hessian analysis to explain why the generator's loss landscape can become ill-conditioned near convergence, especially under the JS-divergence formulation.

7. In the ELBO objective, the KL divergence acts as a regularizer. Show why minimizing $E_{x\sim p_{data}}[KL(q(z|x)\|p(z))]$ encourages consistency of latent codes across different inputs. Use properties of KL and expectations to derive your reasoning.