Basic Knowledge Review

- Matrix derivative: everything comes from the definition of the differential

\[ y(x + dx) = y(x) + A dx + \text{(high order terms)} \]

Derivative, a.k.a Jacobian

In general

\[ y : m \times 1 \quad x : n \times 1 \quad A : m \times n \]

\[ dy = A dx \quad \frac{\partial y}{\partial x} = A \]

Some books use \( A^T \) instead, because they perform the chain rule from right to left
Basic Knowledge Review

- Convex conjugate (Fenchel Conjugate)

for an arbitrary convex function \( f(\cdot) \), there exists a duality function \( g(\cdot) \)

\[
    f(x) = \max_{\lambda} \lambda x - g(\lambda)
\]

\[
    g(\lambda) = \max_x \lambda x - f(x)
\]

Jensen’s equality and convex conjugate plays the key role in approximate inference
Consider a differentiable function \( f: \mathbb{R} \to \mathbb{R} \) that transforms a random variable \( X \) into a random variable \( Y \) by \( Y = f(X) \). Then the pdf of \( Y \) is given by

\[
p(y) = \left| \frac{d}{dy}(f^{-1}(y)) \right| p(f^{-1}(y))
\]

where \( Y = f^{-1}(x) \) with \( \left| \det \frac{\partial^2 x}{\partial y^2} \right| p(x) \).