

# Gradient-Based Optimisation for Variational Inference: A Comparative Study Across Five Bayesian Regression Settings

David Ewing (82171165)  
ENEL445 — Applied Engineering Optimisation

12 May 2026

## Abstract

This report applies the gradient-based optimisation methods from ENEL445 to variational inference across five Bayesian regression settings: linear, quadratic, logistic, hierarchical linear, and hierarchical logistic regression. The unifying objective is maximising the Evidence Lower Bound (ELBO), which converts intractable posterior inference into a tractable optimisation problem.

Four optimisation methods are compared in each setting: Coordinate Ascent Variational Inference (CAVI), gradient ascent with Armijo backtracking, Newton's method with regularised finite-difference Hessian, and BFGS with Wolfe conditions. The variational family is  $q(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with a Cholesky reparameterisation to ensure positive definiteness throughout optimisation.

Reference posteriors are obtained from Gibbs samplers. For the linear and quadratic cases, the conjugate structure admits closed-form Gibbs updates; the variational family is a Normal-Gamma approximation. For the logistic case, the Jaakkola-Jordan (Jaakkola & Jordan, 2000) variational bound restores tractability, and a Pólya-Gamma Gibbs sampler (Polson et al., 2013) provides the reference.

The four test cases form a designed progression in inferential difficulty: the linear case has a conjugate Gaussian likelihood ( $n = 50, p = 2$ ), the quadratic case adds nonlinearity via feature expansion ( $n = 100, p = 3$ ), the logistic case introduces a non-conjugate Bernoulli likelihood ( $n = 200, p = 2$ ), the hierarchical linear case adds group random effects ( $J = 5$  groups,  $N = 150, p = 2$ ), and the hierarchical logistic case combines the Jaakkola-Jordan bound with mixed effects ( $J = 5$  groups,  $N = 150, p = 2$ ). Posterior variance collapse is mild in the linear case, moderate in the quadratic case, minimal in the logistic case, mild in the hierarchical linear case, and symmetric (both coefficients) in the hierarchical logistic case. CAVI is the most efficient method across all five settings. BFGS is the recommended gradient-based alternative when closed-form CAVI updates are not available.

Awaiting review by Professor Le Yang