Thermodynamic Variational Laplace (ThermoVL): A concise derivation and algorithmic specification

Dr Alexander D. Shaw
a.d.shaw@exeter.ac.uk
https://cpnslab.com

October 8, 2025

Abstract

This note derives and specifies a thermodynamically–tempered Variational Laplace (VL) procedure matching the implementation in `fitVariationalLaplaceThermo.m`. We state the probabilistic model, define the tempered objective, give closed–form expressions for the gradient, curvature, Gaussian variational posterior, and the variational free energy (VFE), and present a practical Newton–Laplace update with backtracking. A short mapping between symbols and code variables is provided for cross–reference.

1 Model and notation

Let $y \in \mathbb{R}^n$ denote observed data and $g(\theta) \in \mathbb{R}^n$ a (possibly nonlinear) forward model with parameters $\theta \in \mathbb{R}^p$. We assume additive Gaussian noise

$$y = g(\theta) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \Sigma_y), \tag{1}$$

with precision $\Lambda_y := \Sigma_y^{-1}$. A Gaussian prior on parameters is

$$\theta \sim \mathcal{N}(\mu_0, \Sigma_0), \qquad \Lambda_0 := \Sigma_0^{-1}. \tag{2}$$

When $\Sigma_y = \sigma^2 \mathbf{I}$ the precision is $\Lambda_y = \sigma^{-2}\mathbf{I}$, but the derivations below hold for general (fixed) Λ_y.

Log-likelihood and log-prior. Define the log terms (up to additive constants independent of θ):

$$\ell(\theta) := \ln p(y \mid \theta) = -\frac{1}{2}\,(y - g(\theta))^{\mathsf{T}} \Lambda_y (y - g(\theta)), \tag{3}$$

$$\pi(\theta) := \ln p(\theta) = -\frac{1}{2}\,(\theta - \mu_0)^{\mathsf{T}} \Lambda_0 (\theta - \mu_0). \tag{4}$$

2 Thermodynamic tempering

Introduce an inverse-temperature (or annealing) parameter $\beta \in [0, 1]$ that *tempers* the likelihood:

$$\ln p_\beta(y, \theta) := \beta\,\ell(\theta) + \pi(\theta) + \text{const.} \tag{5}$$

For $\beta = 1$ we recover the standard (untempered) objective. Homotopy/annealing schemes typically start from small β and increase to 1 to improve convergence for highly non-linear posteriors.

3 Variational Laplace approximation

We approximate $p(\theta \mid y)$ by a Gaussian $q(\theta) = \mathcal{N}(\mu, \Sigma)$ and maximise the variational free energy (VFE):

$$F[q] := \mathbb{E}_q[\ln p_\beta(y, \theta)] - \mathbb{E}_q[\ln q(\theta)], \tag{6}$$

which is a lower bound on $\ln p_\beta(y)$ and is tight at the Laplace point.

3.1 Local quadratic expansion (Laplace)

Let

$$\mathcal{L}_\beta(\theta) := \beta\, \ell(\theta) + \pi(\theta). \tag{7}$$

Around the current mean μ, take a second-order Taylor expansion:

$$\mathcal{L}_\beta(\theta) \approx \mathcal{L}_\beta(\mu) + g^\top(\theta - \mu) - \tfrac{1}{2}(\theta - \mu)^\top H\,(\theta - \mu), \tag{8}$$

where the *score* and (negative) Hessian at μ are

$$g := \nabla_\theta \mathcal{L}_\beta(\mu) = \beta\, \nabla_\theta \ell(\mu) + \nabla_\theta \pi(\mu), \tag{9}$$

$$H := -\nabla_\theta^2 \mathcal{L}_\beta(\mu) = -\beta\, \nabla_\theta^2 \ell(\mu) - \nabla_\theta^2 \pi(\mu). \tag{10}$$

For the Gaussian prior we have $\nabla_\theta \pi(\mu) = -\Lambda_0(\mu - \mu_0)$ and $-\nabla_\theta^2 \pi(\mu) = \Lambda_0$.

Likelihood derivatives via Gauss–Newton. Let $r(\theta) := y - g(\theta)$ and $J(\theta) := \partial g(\theta)/\partial\theta \in \mathbb{R}^{n\times p}$. A Gauss–Newton approximation yields

$$\nabla_\theta \ell(\mu) \approx J(\mu)^\top \Lambda_y\, r(\mu), \tag{11}$$

$$-\nabla_\theta^2 \ell(\mu) \approx J(\mu)^\top \Lambda_y\, J(\mu), \tag{12}$$

neglecting second derivatives of g. This is standard in VL/DCM and ensures a positive semi-definite curvature from the likelihood term.

3.2 Gaussian variational posterior

The Laplace posterior is

$$\Sigma \;=\; H^{-1}, \qquad \mu \leftarrow \mu + \Sigma\, g, \tag{13}$$

optionally with a line-search/backtracking step length $\eta \in (0, 1]$ so that $\mu \leftarrow \mu + \eta\, \Sigma g$ ensures monotone F-increase.

With Gauss–Newton,

$$H \;\approx\; \beta\, J(\mu)^\top \Lambda_y J(\mu) \;+\; \Lambda_0. \tag{14}$$

3.3 Variational free energy at the Laplace point

Up to an additive constant independent of (μ, Σ),

$$F(\mu, \Sigma) \;=\; \mathcal{L}_\beta(\mu) \;+\; \tfrac{1}{2} \log\det(2\pi e\, \Sigma). \tag{15}$$

Equivalently, using $H = \Sigma^{-1}$,

$$F(\mu, \Sigma) \;=\; \beta\, \ell(\mu) + \pi(\mu) \;-\; \tfrac{1}{2} \log\det(H) \;+\; \text{const.} \tag{16}$$

4 Algorithm (Newton–Laplace with tempering and backtracking)

Algorithm 1 Thermodynamic Variational Laplace (ThermoVL)

Require: data y, forward model $g(\cdot)$, Jacobian $J(\cdot)$, prior (μ_0, Σ_0), precision Λ_y, schedule $\{\beta_t\}_{t=1}^{T}$, tolerances

Initialise $\mu \leftarrow \mu_0$, $\Lambda_0 \leftarrow \Sigma_0^{-1}$
for $t = 1, \ldots, T$ **do** ▷ anneal β up to 1
    $\beta \leftarrow \beta_t$
    repeat
        $r \leftarrow y - g(\mu); \quad J \leftarrow J(\mu)$
        $g \leftarrow \beta\, J^{\mathsf{T}} \Lambda_y r \; - \; \Lambda_0(\mu - \mu_0)$ ▷ score, (9)
        $H \leftarrow \beta\, J^{\mathsf{T}} \Lambda_y J \; + \; \Lambda_0$ ▷ curvature, (14)
        Solve $H\,\Delta\theta = g$ (e.g. Cholesky); set candidate $\mu^\star \leftarrow \mu + \eta\,\Delta\theta$
        Compute $F^\star := F(\mu^\star, H^{-1})$ via (15)
        if $F^\star \geq F(\mu, H^{-1})$ **then**
            Accept: $\mu \leftarrow \mu^\star$
        else
            Backtrack: reduce $\eta \in (0,1)$ and retry
        end if
    until converged in ELBO/VFE or step size
end for
return μ, $\Sigma \leftarrow H^{-1}$, F, diagnostics

5 Hyperparameters and noise

If a noise variance σ^2 (or a set of precision hyperparameters) is inferred, place a prior on σ^{-2} (e.g. Gamma) and alternate updates (or embed them in F). With a single homoscedastic variance,

$$\hat{\sigma}^2 \;=\; \frac{1}{n}\,\mathbb{E}_q\Big[\|y - g(\theta)\|^2_{\Lambda_y}\Big] \;\approx\; \frac{1}{n}\,\|y - g(\mu)\|^2_{\Lambda_y} \;+\; \frac{1}{n}\,\mathrm{tr}\big(J(\mu)^{\mathsf{T}}\Lambda_y J(\mu)\,\Sigma\big). \tag{17}$$

More general heteroscedastic or structured noise can be handled by parameterising Λ_y and including its derivatives in g, H.

6 Connections and remarks

- Setting $\beta = 1$ recovers standard VL. The Gauss–Newton Hessian (12) is widely used in DCM/DEM for stability.
- The tempering path $\{\beta_t\}$ implements a homotopy from prior to full posterior, facilitating convergence for non-convex objectives.
- Backtracking/damping ensures monotonic VFE ascent and guards against overshooting in highly non-linear regimes.
- Low-rank or structured approximations to H^{-1} can be used to reduce computational cost in high dimensions.

7 Symbol–to–code map (minimal)

Math	Meaning	Typical code symbol in `fitVariationalLaplaceThermo.m`
y	data (vector)	`y`
$g(\theta)$	forward model	`f` / function handle
$J(\mu)$	Jacobian of g at μ	`J` (via user/model Jacobian)
μ	posterior mean	`m`
Σ	posterior covariance	`V`
μ_0, Σ_0	prior mean/covariance	`m0, S0`
Λ_0	prior precision	`inv(S0)` or explicit
Λ_y	noise precision	`inv(SigmaY)` or scalar precision
β	tempering weight	`beta` / annealing schedule
F	variational free energy	`F`

Reproducibility. For maximal fidelity to the code, log the accepted step size η, β schedule, convergence tolerances, Cholesky failures (if any), and the VFE trajectory $F^{(k)}$.

Acknowledgements. This write-up is designed to accompany and document `fitVariationalLaplaceThermo` Please cite appropriately if reused.