Non-Symmetric Information Operators in GMRFs for PDE Inference

Yumeng Shi YUMENG.SHI@ALUMNOS.UPM.ES
Computational Intelligence Group
Universidad Politécnica de Madrid
Madrid, Spain

Concha Bielza MCBIELZA@FI.UPM.ES
Computational Intelligence Group
Universidad Politécnica de Madrid
Madrid, Spain

Pedro Larrañaga PEDRO.LARRANAGA@FI.UPM.ES
Computational Intelligence Group
Universidad Politécnica de Madrid
Madrid, Spain

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Probabilistic approaches to partial differential equations (PDEs) cast the solution as a Bayesian inference problem, in which the unknown field is inferred from a prior distribution and a likelihood derived from the governing equations. When the prior is Gaussian-Markov, the solution admits a Gaussian Markov Random Field (GMRF) representation with a sparse precision matrix, enabling efficient inference and marginal uncertainty estimates. In this framework, physical constraints enter through information operators that define the likelihood. Standard formulations rely on symmetric operators, which encode a non-directional flow of information. In advection-dominated regimes, this mismatch produces spurious oscillations and unreliable uncertainty estimates. In this work we introduce a non-symmetric information operator that embeds a physically motivated directional preference into the likelihood. Although the operator is non-symmetric, the resulting posterior precision matrix remains symmetric and positive definite, preserving full compatibility with standard sparse GMRF inference. We evaluate the method on the viscous Burgers' equation across diffusion- to advection-dominated regimes. The non-symmetric operator suppresses spurious oscillations and matches finite-element accuracy. Its posterior uncertainty remains reliable in advection-dominated regimes where the symmetric operator becomes overconfident.

Keywords: Gaussian Markov Random Fields, Non-Symmetric Information Operators, Probabilistic Numerics, Uncertainty Quantification, Advection-Dominated PDEs

1. Introduction

Partial differential equations (PDEs) provide a fundamental framework for modeling physical systems across science and engineering (Karniadakis et al., 2021). In many practical settings, however, observations of these systems are sparse, noisy, and incomplete, making it challenging to reliably infer the underlying physical field. A useful solution must therefore do more than estimate the field: it must also convey how far that estimate can be trusted.

© 2026 Y. Shi, C. Bielza & P. Larrañaga.

A confident but wrong answer is often more dangerous than one openly flagged as uncertain. A principled approach to both challenges requires a probabilistic inference framework capable of integrating observational data with governing physical laws, while representing the uncertainty that remains in the inferred solution (Chen et al., 2021; Hamelijnck et al., 2024).

Several representations can support this kind of inference. Physics-informed neural networks equipped with Bayesian training can address both forward and inverse problems. However, their uncertainty estimates depend on priors or dropout rates that cannot be principled without additional information, and no closed-form posterior is generally available (Yang et al., 2021; Shih et al., 2025; Shi et al., 2025). Classical finite element methods equipped with uncertainty quantification, such as polynomial chaos expansion or Monte Carlo propagation, retain strong convergence guarantees, but their computational cost grows rapidly with the dimension of the input uncertainty (Xiu and Karniadakis, 2002; Lüthen et al., 2021; Kumar et al., 2022). Gaussian Markov random fields (GMRFs) offer a different trade-off (Lindgren et al., 2011, 2024). Because both the prior and the likelihood are Gaussian by construction, the posterior is available in closed form, with an explicit mean and covariance, while the sparsity of the precision matrix preserves scalability to high-dimensional domains.

We adopt this representation explicitly: the unknown PDE solution is treated as a latent random field, allowing observational data and physical constraints to be unified within a Bayesian inference framework (Hennig et al., 2022; Pförtner et al., 2022). The sparsity of the precision matrix arises from exploiting local conditional independence structure over the discretized domain, a structure that full Gaussian processes cannot exploit at scale. Physical constraints enter through likelihood terms called information operators (Cockayne et al., 2019), which map the latent field to residual observations of the governing equations and thereby determine how these constraints propagate as dependencies across the Markov graph.

Existing approaches predominantly rely on symmetric information operators, which induce symmetric interactions in the likelihood construction (McGreivy and Hakim, 2024). This symmetry imposes a non-directional flow of information across the graph, treating the influence between neighboring nodes as mutually reciprocal regardless of the underlying physical dynamics. In graphical terms, a symmetric information operator encodes an undirected dependency structure, in which the conditional influence between two nodes is symmetric by construction. While such an assumption is adequate for diffusion-dominated systems, where the interactions are genuinely reciprocal, it becomes fundamentally mismatched in the presence of directional dependencies. Advection is inherently directed: information along a characteristic propagates downstream but not upstream, a non-reciprocal dependency that an undirected, symmetric construction cannot represent, see Section 2.1. As a result, symmetric constructions are unable to faithfully represent the underlying physics, leading to systematic artifacts such as spurious oscillations near regions with sharp gradients (Michel et al., 2021).

To address this limitation, we propose a non-symmetric information operator that introduces a directional preference into the inference process. We develop and validate this construction in the context of advection–diffusion equations, using the viscous Burgers' equation across multiple viscosity regimes as a benchmark, whose self-steepening, near-

shock gradients typify the core challenge of advection-dominated transport. Our construction thus injects a directed, non-reciprocal dependency into the likelihood while the resulting posterior precision matrix remains that of an undirected GMRF, expressing directional information without sacrificing the sparse, symmetric structure that makes GMRF inference efficient.

The main contributions of this work are as follows: (i) Introduce a non-symmetric information operator that encodes directed, non-reciprocal dependencies within a GMRF; (ii) Show that this non-symmetric construction nonetheless yields a posterior precision matrix that is symmetric, positive definite, and sparse, so that standard GMRF inference applies unchanged; (iii) Show that the resulting posterior uncertainty remains reliable in advection-dominated regimes, whereas the symmetric formulation becomes overconfident.

The remainder of this paper is organized as follows. Section 2 provides the necessary background. Section 3 presents the proposed method. Section 4 reports on experimental results. Section 5 concludes with a discussion of limitations and future directions.

2. Background

2.1. Diffusion, advection and advection-dominated regimes

Many physical models involve two elementary mechanisms that can appear separately or together. Diffusion spreads a field without a preferred direction: a local disturbance relaxes symmetrically, with the influence between two points being mutually reciprocal, as in the heat equation or the Poisson equation. Advection transports a field along a velocity: the field is carried in a preferred direction set by the flow, rather than spreading evenly. A canonical model combining both is

$$\frac{\partial u}{\partial t} + \underbrace{\boldsymbol{a} \cdot \nabla u}_{\text{advection}} = \underbrace{\nu \nabla^2 u}_{\text{diffusion}}, \tag{1}$$

where $u(\mathbf{x}, t)$ is the field of interest, $\boldsymbol{a}$ is the advection velocity, and $\nu \geq 0$ is the viscosity. The first-order term is directional, reflecting the orientation of $\boldsymbol{a}$, while the second-order term is rotationally symmetric.

When ν is small relative to $|\boldsymbol{a}|$, the dynamics are advection-dominated: directional transport dominates diffusive smoothing, and the field develops steep, near-discontinuous gradients that are carried downstream rather than relaxing symmetrically. This is in contrast to diffusion-dominated regimes, where the response to a local disturbance remains essentially reciprocal. Advection-dominated regimes are known to be difficult to model numerically: representations that treat all directions alike are blind to the direction of transport and tend to produce spurious oscillations. In graphical terms, this is a statement about edge direction. Diffusion is reciprocal and maps naturally to undirected edges, whereas advection—information travels downstream but not upstream—calls for directed ones. A symmetric information operator can only produce undirected edges; our method supplies the missing direction at the operator level while keeping the posterior graph an ordinary undirected GMRF.

Burgers' equation. Our experiments focus on the viscous Burgers' equation,

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial x} = \nu\,\frac{\partial^2 u}{\partial x^2}, \tag{2}$$

obtained by letting the advection velocity equal the field itself. It is a standard benchmark for advection-dominated behavior: small ν produces self-steepening, near-shock gradients, while large ν yields smooth, diffusion-dominated behavior. We evaluate across multiple viscosity regimes to probe both extremes.

2.2. Probabilistic inference for PDEs via Gaussian Markov random fields

Section 2.1 framed the advection problem as a physical mismatch; resolving it requires a computational framework that turns the governing equations and observational data into a tractable posterior over the unknown field. Probabilistic PDE solvers (Hennig et al., 2022) provide such a framework, casting the computation of a solution as Bayesian inference on a latent random field. The standard instantiation places a Gaussian process (GP) prior over the solution and conditions on the initial condition, boundary data, and PDE constraints (Pförtner et al., 2022); for linear PDEs the posterior is then available in closed form. Its cost, however, is governed by a dense covariance matrix, so inference scales as $\mathcal{O}(N^3)$ in the number of discretization points and becomes prohibitive on the fine grids that advection-dominated regimes demand.

A Gaussian Markov random field (GMRF) replaces the dense covariance with a sparse precision matrix (Lindgren et al., 2024),

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}),$$

where $\boldsymbol{Q}$ encodes the conditional independence structure of the discretized field: its sparsity pattern mirrors the neighborhood of each node, and inference reduces to sparse Cholesky factorization at a cost far below the cubic scaling of GP inference. We adopt this representation as a space–time trajectory, in which a linear advection–diffusion evolution operator couples successive time layers and the Markov property in time renders $\boldsymbol{Q}$ block-tridiagonal (Weiland et al., 2025).

Physical constraints enter this framework through the likelihood, via the information operators introduced earlier, which map the latent field to PDE residuals. The choice of operator is the single modeling decision that fixes the dependency structure of the posterior graph. Standard formulations take it to be symmetric, which yields a sparse, well-conditioned posterior but encodes the non-directional coupling identified in Section 2.1, the very coupling that misrepresents advection. Section 3 modifies exactly this choice.

3. Method

3.1. Probabilistic PDE Formulation

The unknown solution is modeled as a latent Gaussian random field,

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}), \tag{3}$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the prior mean (typically set to zero in practice) and N is the number of discretization nodes over the space–time domain.

The prior precision $\boldsymbol{Q}$ is constructed from a linear stochastic PDE (SPDE) (Lindgren et al., 2011) that captures the diffusion and mean advection of the governing dynamics. Introducing a stochastic forcing term $\mathcal{W}$ scaled by a prior-strength parameter $\tau_p \in \mathbb{R}$, we obtain the advection–diffusion SPDE

$$\frac{\partial u}{\partial t} + \bar{c}\frac{\partial u}{\partial x} - \nu\frac{\partial^2 u}{\partial x^2} = \tau_p \mathcal{W}, \tag{4}$$

where $\bar{c}$ is the mean advection velocity.

Discretizing this SPDE yields the sparse precision matrix $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$, whose sparsity pattern encodes the conditional independence structure of the latent field over the computational nodes.

The prior $\boldsymbol{Q}$ encodes the mean dynamics but not yet the PDE constraints themselves; these enter through the likelihood. We treat the PDE residuals as probabilistic observations of the latent field, giving a likelihood model of the form

$$\boldsymbol{y} \mid \boldsymbol{u} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{u} + \boldsymbol{b}, \boldsymbol{Q}_\epsilon^{-1}), \tag{5}$$

where $\boldsymbol{y} \in \mathbb{R}^M$ collects the PDE residuals evaluated at M collocation points, $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is the information operator mapping the latent field to PDE residuals, $\boldsymbol{b} \in \mathbb{R}^M$ encodes the source terms, and $\boldsymbol{Q}_\epsilon \in \mathbb{R}^{M\times M}$ is block-diagonal, assigning independent precision weights to the PDE-residual, initial-condition, and boundary-condition observations, respectively.

$$\boldsymbol{Q}_\epsilon = \mathrm{blockdiag}\big(\lambda_{\mathrm{pde}}\boldsymbol{I},\ \lambda_{\mathrm{ic}}\boldsymbol{I},\ \lambda_{\mathrm{bnd}}\boldsymbol{I}\big). \tag{6}$$

The precision weights in $\boldsymbol{Q}_\epsilon$ control how strictly the constraints are enforced: taking $\boldsymbol{Q}_\epsilon = \lambda\boldsymbol{I}$ with $\lambda \gg 1$ approximates their deterministic enforcement.

Conditioning the Gaussian prior on this Gaussian likelihood yields a Gaussian posterior $\boldsymbol{u} \mid \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{post}}, \boldsymbol{Q}_{\mathrm{post}}^{-1})$, with posterior precision matrix

$$\boldsymbol{Q}_{\mathrm{post}} = \boldsymbol{Q} + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A}, \tag{7}$$

and posterior mean

$$\boldsymbol{\mu}_{\mathrm{post}} = \boldsymbol{\mu} + \boldsymbol{Q}_{\mathrm{post}}^{-1}\boldsymbol{A}^\top \boldsymbol{Q}_\epsilon(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} - \boldsymbol{b}). \tag{8}$$

Here the choice of λ trades off constraint satisfaction against the conditioning of the posterior precision matrix $\boldsymbol{Q}_{\mathrm{post}}$: larger λ enforces the constraints more strictly but worsens its conditioning.

When the PDE is nonlinear, so is the likelihood in $\boldsymbol{u}$. Following Weiland et al. (2025), we linearize it with a Gauss–Newton/Laplace scheme: from an initial estimate $\boldsymbol{\mu}^{(0)}$, each iteration linearizes the observation model around the current estimate $\boldsymbol{u}^{(k)}$ via the Jacobian $\boldsymbol{J}^{(k)} = \nabla_{\boldsymbol{u}} f(\boldsymbol{u}^{(k)})$ and updates the precision and mean:

$$\boldsymbol{Q}^{(k)} = \boldsymbol{Q} + \boldsymbol{J}^{(k)\top}\boldsymbol{Q}_\epsilon \boldsymbol{J}^{(k)}, \tag{9}$$

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} + \Big(\boldsymbol{Q}^{(k)}\Big)^{-1}\boldsymbol{J}^{(k)\top}\boldsymbol{Q}_\epsilon\Big(\boldsymbol{y} - f\big(\boldsymbol{\mu}^{(k)}\big)\Big). \tag{10}$$

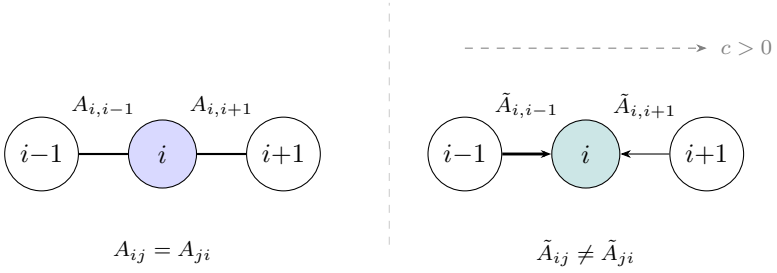


Figure 1: Symmetric (left) vs. non-symmetric (right) information operators on a chain of nodes. The symmetric construction gives undirected edges ($A_{ij} = A_{ji}$); the proposed construction breaks this symmetry ($\tilde{A}_{ij} \neq \tilde{A}_{ji}$), yielding edges directed along the flow c.

3.2. Non-symmetric Information Operator

In the above formulation, the information operator $\boldsymbol{A}$ determines how PDE constraints are encoded as edges in the posterior graphical model: each row of $\boldsymbol{A}$ corresponds to a local constraint that introduces dependencies between neighboring nodes in the graph.

Standard constructions define $\boldsymbol{A}$ through the inner product of test functions $\{\phi_i\}$ with the differential operator $\mathcal{L}$:

$$(\boldsymbol{A}\boldsymbol{u})_i = \langle \phi_i, \mathcal{L}u \rangle. \tag{11}$$

For the self-adjoint (diffusion) part of the operator, this construction yields a symmetric matrix and, in graph terms, undirected dependencies: node i influences each neighbor. The directional character of advection, by contrast, is not captured by such symmetric couplings, which motivates a construction that can encode asymmetric edges. Figure 1 illustrates this contrast for a one-dimensional chain of nodes.

This motivates modifying the test functions so that $\boldsymbol{A}$ can encode the asymmetric, direction-dependent edges shown in Figure 1. We present the construction for a general advection velocity $\boldsymbol{c}$; in our one-dimensional experiments it reduces to a scalar, which $\boldsymbol{c} \cdot \nabla$ becoming $c\partial_x$.

To encode a directional preference, we bias each test function along the flow, replacing ϕ_i with

$$\psi_i = \phi_i + \tau_i \, c_i \, \partial_x \phi_i, \tag{12}$$

where c_i is the local advection velocity (with $c_i = u_i$ for Burgers' equation) and $\tau_i \in \mathbb{R}$ is the stabilization parameter at node i, specified below. The added term $\tau_i \, c_i \, \partial_x \phi_i$ tilts the test function upwind, so that the residual at node i is weighted more heavily against the direction of transport.

Substituting ψ_i into the operator yields its non-symmetric counterpart,

$$(\tilde{\boldsymbol{A}}\boldsymbol{u})_i = \langle \psi_i, \mathcal{L}u \rangle = \langle \phi_i, \mathcal{L}u \rangle + \tau_i \langle c_i \, \partial_x \phi_i, \mathcal{L}u \rangle. \tag{13}$$

Expanding in the basis $u(x) \approx \sum_{j=1}^{N} u_j \, \phi_j(x)$ separates each entry of $\tilde{\boldsymbol{A}} \in \mathbb{R}^{N \times N}$ into the original symmetric part and a directional correction,

$$\tilde{A}_{ij} = \underbrace{\langle \phi_i, \, \mathcal{L}\phi_j \rangle}_{A_{ij}} + \tau_i \underbrace{\langle c_i \, \partial_x \phi_i, \, \mathcal{L}\phi_j \rangle}_{S_{ij}}, \tag{14}$$

Algorithm 1 Non-symmetric GMRF PDE solver

Input: Prior $(\boldsymbol{\mu}, \boldsymbol{Q})$, PDE residual f, observations $\boldsymbol{y}$, observation precision $\boldsymbol{Q}_\epsilon$
Output: Posterior mean $\boldsymbol{\mu}^*$ and marginal variances $\boldsymbol{\sigma}^2$

$\boldsymbol{\mu}^{(0)} \leftarrow \boldsymbol{\mu}$ **for** $k = 0, 1, \ldots$ **do**
  $\boldsymbol{\tau}^{(k)} \leftarrow \boldsymbol{\tau}(\boldsymbol{\mu}^{(k)})$; // stabilization at current estimate
  $\tilde{\boldsymbol{A}}^{(k)} \leftarrow \boldsymbol{A} + \mathrm{diag}(\boldsymbol{\tau}^{(k)})\, \boldsymbol{S}$; // assemble non-symmetric operator
  $\boldsymbol{J}^{(k)} \leftarrow \nabla_{\boldsymbol{u}} f(\boldsymbol{\mu}^{(k)})$; // Jacobian, via automatic differentiation
  $\boldsymbol{Q}^{(k)} \leftarrow \boldsymbol{Q} + \boldsymbol{J}^{(k)\top} \boldsymbol{Q}_\epsilon \boldsymbol{J}^{(k)}$; // posterior precision (sparse, SPD)
  Solve $\boldsymbol{Q}^{(k)} \boldsymbol{\delta} = \boldsymbol{J}^{(k)\top} \boldsymbol{Q}_\epsilon \big(\boldsymbol{y} - f(\boldsymbol{\mu}^{(k)})\big)$; // sparse Cholesky solve
  $\boldsymbol{\mu}^{(k+1)} \leftarrow \boldsymbol{\mu}^{(k)} + \boldsymbol{\delta}$ **if** *converged* **then**
    break
  end
end
$\boldsymbol{Q}^* \leftarrow \boldsymbol{Q}^{(k)}$; // Laplace posterior precision
return $\boldsymbol{\mu}^* = \boldsymbol{\mu}^{(k)}$ *and* $\boldsymbol{\sigma}^2 = \mathrm{diag}\big((\boldsymbol{Q}^*)^{-1}\big)$; // selected inversion

or, in matrix form,

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \mathrm{diag}(\boldsymbol{\tau})\, \boldsymbol{S}, \tag{15}$$

where $\boldsymbol{S}$ is the directional correction matrix and $\mathrm{diag}(\boldsymbol{\tau})$ collects the nodewise stabilization parameters. This single term $\mathrm{diag}(\boldsymbol{\tau})\boldsymbol{S}$ is the whole modification: it breaks the symmetry of $\boldsymbol{A}$ ($\tilde{A}_{ij} \neq \tilde{A}_{ji}$), turning the undirected edges of Figure 1 into directed ones aligned with the flow. Because $\boldsymbol{S}$ shares the neighborhood structure of $\boldsymbol{A}$, the operator $\tilde{\boldsymbol{A}}$ remains sparse.

It remains to specify the stabilization parameter τ_i. Because the backward-Euler time coupling is assembled jointly across all time layers and solved as a single space–time system, rather than sequentially layer by layer, τ_i must reflect the local advective and diffusive scales together with the time step Δt of the backward-Euler discretization. Following the algebraic multiscale form standard for transient advection–diffusion problems (Shakib et al., 1991; Tezduyar and Osawa, 2000; Codina, 2002), we compute

$$\tau_i = \left[\left(\frac{2}{\Delta t} \right)^2 + \left(\frac{2\|c_i\|}{h_i} \right)^2 + \left(\frac{4\nu}{h_i^2} \right)^2 \right]^{-1/2}, \tag{16}$$

where h_i is the local element size and $\|c_i\|$ the local advection speed. The transient term bounds $\tau_i \leq \Delta t/2$, which keeps τ_i well-defined where the advection speed $c = u$ becomes small, as it does near the shock in Burgers' equation.

Since $c_i = u_i$ for Burgers' equation, τ_i depends on the latent field itself and must be re-evaluated at every Gauss–Newton iteration from the current estimate $u^{(k)}$. The Jacobian $\boldsymbol{J}^{(k)}$ of Section 3.1 must therefore account for this dependence, including the term $\partial \tau_i / \partial u$; omitting it yields an inexact linearization and biases the Gauss–Newton step. We obtain this term by forward-mode automatic differentiation.

For a linear PDE the residual is affine, $f(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{b}$, a single iteration is then exact, and Algorithm 1 reduces to the closed-form posterior of Eqs. (7)–(8).

3.3. Posterior Structure and Inference

Replacing $\boldsymbol{A}$ with the non-symmetric operator $\tilde{\boldsymbol{A}}$ of Section 3.2 leaves the posterior of Section 3 unchanged in form:

$$\boldsymbol{Q}_{\text{post}} = \boldsymbol{Q} + \tilde{\boldsymbol{A}}^{\top} \boldsymbol{Q}_{\epsilon} \tilde{\boldsymbol{A}}. \tag{17}$$

The asymmetry of $\tilde{\boldsymbol{A}}$ does not propagate to $\boldsymbol{Q}_{\text{post}}$: the quadratic form $\tilde{\boldsymbol{A}}^{\top} \boldsymbol{Q}_{\epsilon} \tilde{\boldsymbol{A}}$ is symmetric positive-semidefinite for any $\tilde{\boldsymbol{A}}$, so $\boldsymbol{Q}_{\text{post}}$ remains symmetric and positive definite (Appendix A). Since $\boldsymbol{S}$ shares the neighborhood structure of $\boldsymbol{A}$, it also adds no fill-in. The posterior precision is therefore an ordinary sparse SPD GMRF precision, and the inference of Algorithm 1 applies without modification.

Beyond the point estimate $\boldsymbol{\mu}_{\text{post}}$, the same posterior supplies its own uncertainty. The marginal variance $\boldsymbol{\sigma}^2$,

$$\boldsymbol{\sigma}^2 = \boldsymbol{Q}_{\text{post}}^{-1}, \tag{18}$$

which we obtain without forming the dense inverse via selected inversion of the sparse Cholesky factor of $\boldsymbol{Q}_{\text{post}}$. These marginal standard deviations σ_i define the per-node credible intervals $\mu_{\text{post},i} \pm 1.96\,\sigma_i$ used to assess calibration in Section 4.3.

4. Experiments

4.1. Experimental Setup

We evaluate the proposed method on Burgers' equation,

$$\partial_t u + u\,\partial_x u = \nu\,\partial_{xx} u, \qquad (x,t) \in [-1,1] \times [0,1], \tag{19}$$

with homogeneous Dirichlet boundary conditions $u(\pm 1, t) = 0$ and initial condition $u(x,0) = -\sin(\pi x)$. We vary the viscosity $\nu \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ to span diffusion-dominated and advection-dominated regimes. The spatiotemporal domain is discretized using 200 nodes uniformly spaced ($h = 10^{-2}$) and a uniform time step of $\Delta t = 0.005$. PDE constraints are imposed at every node and time level through $\boldsymbol{Q}_{\epsilon}$, with $\lambda_{\text{pde}} = 10^6$ and $\lambda_{\text{ic}} = \lambda_{\text{bnd}} = 10^8$.

Baselines. We compare three methods throughout all experiments:

- **Reference:** a finite-element solution on a grid eight times finer than the inference grid ($h_{\text{ref}} = h/8$), grid-converged to below 10^{-6} at the most demanding case $\nu = 10^{-5}$.
- **Symmetric operator**: our GMRF-based probabilistic solver with a symmetric information operator, i.e. $\tilde{\boldsymbol{A}} = \boldsymbol{A}$ (corresponding to $\boldsymbol{\tau} = 0$).
- **Non-symmetric operator**: the proposed method with the non-symmetric information operator $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \operatorname{diag}(\boldsymbol{\tau})\,\boldsymbol{S}$, as defined in Eq. (15).

The two GMRF variants share the same prior, observation precisions, discretization, and linearization scheme; the information operator is the only difference between them, isolating the effect of the operator itself. Errors are computed relative to the reference solution; no error is reported for the reference itself.

We report two complementary metrics. Accuracy is measured by the ℓ_2 error,

$$\ell_2(u) = \left(\frac{1}{N} \sum_{i=1}^{N} (u_i - u_i^*)^2 \right)^{1/2}, \tag{20}$$

where u^* denotes the reference solution, and the maximum error $L^\infty(u) = \max_i |u_i - u_i^*|$, which is sensitive to localized errors at the shock.

We assess spurious oscillation using two complementary measures computed on the posterior mean at $t = 1$: the total variation

$$\mathrm{TV}(u) = \sum_i |u_{i+1} - u_i|, \tag{21}$$

and the sharpness ratio $\max_i |u_{i+1} - u_i| \,/\, \max_i |u_{i+1}^* - u_i^*|$, the largest single-node jump in the posterior mean relative to that of the reference. A value near 1 indicates the shock is resolved at the correct steepness, while values above (below) 1 indicate spurious sharpening (excessive smoothing).

The total variation admits a physical upper bound. Since $\mathrm{TV}(u_0) = 4$ for $u_0 = -\sin(\pi x)$ and the total variation of the viscous Burgers' equation is non-increasing in time, the exact solution satisfies $\mathrm{TV}(u(t)) \leq 4$ for all t. A posterior mean whose total variation far exceeds this ceiling must therefore contain non-physical fluctuations. The sharpness ratio complements this bound: total variation flags oscillation globally, while sharpness checks whether the shock itself is resolved at the correct steepness.

4.2. Accuracy and Oscillation Suppression

We analyze posterior stability across the viscosity range, taking the converged finite-element reference of Section 4.1 as ground truth. As ν decreases, the reference develops an increasingly sharp front at $x = 0$ by $t = 1$, and the symmetric and non-symmetric operators diverge accordingly.

At $\nu = 10^{-2}$ the two operators are indistinguishable (Table 1), both reaching $\ell_2 \approx 9.5 \times 10^{-3}$: diffusion still dominates at the grid scale, so the symmetric likelihood is not yet misspecified. As ν decreases, the reference develops a sharp front and the two operators separate sharply (Figure 2): the symmetric operator develops grid-scale oscillations that grow as $\nu \to 0$, while the non-symmetric operator tracks the reference throughout.

Table 1 quantifies shows the symmetric total variation climbs to 243.438 at $\nu = 10^{-5}$—nearly two orders of magnitude above the reference and far past the physical ceiling $TV \leq 4$—confirming that the oscillations are non-physical. The non-symmetric total variation stays near the reference (2.9–3.2) at every viscosity, and its ℓ_2 error remains below 10^{-2}, a factor of 96 lower than the symmetric operator at $\nu = 10^{-5}$.

One might object that the non-symmetric operator suppresses oscillation only by smearing the shock. The sharpness ratio rules this out. The non-symmetric mean's largest jump stays within 10% of the reference at every viscosity (sharpness 0.98–1.10): the front is resolved at the correct steepness, not smoothed. The symmetric operator instead reaches a sharpness of 2.57 at $\nu = 10^{-5}$—its largest jump is more than twice that of the reference—not because the shock is sharper, but because of the oscillatory ringing visible in Figure 2.

Table 1: Accuracy and total variation at $t = 1$. $\mathrm{TV}_{\mathrm{ref}}$ is the reference value; sharpness is the largest jump between adjacent nodes relative to the reference.

ν	Method	ℓ_2	L^∞	TV	$\mathrm{TV}_{\mathrm{ref}}$	Sharpness
10^{-2}	Symmetric	9.45×10^{-3}	1.94×10^{-2}	2.696	2.683	1.00
	Non-symmetric	9.57×10^{-3}	1.94×10^{-2}	2.681		0.99
10^{-3}	Symmetric	3.49×10^{-2}	2.75×10^{-1}	5.481	2.904	1.38
	Non-symmetric	2.08×10^{-3}	1.65×10^{-2}	2.878		0.98
10^{-4}	Symmetric	1.78×10^{-1}	6.44×10^{-1}	42.510	2.910	1.67
	Non-symmetric	5.93×10^{-3}	5.91×10^{-2}	3.148		1.08
10^{-5}	Symmetric	6.68×10^{-1}	1.09×10^{0}	243.438	2.911	2.57
	Non-symmetric	6.98×10^{-3}	6.96×10^{-2}	3.189		1.10

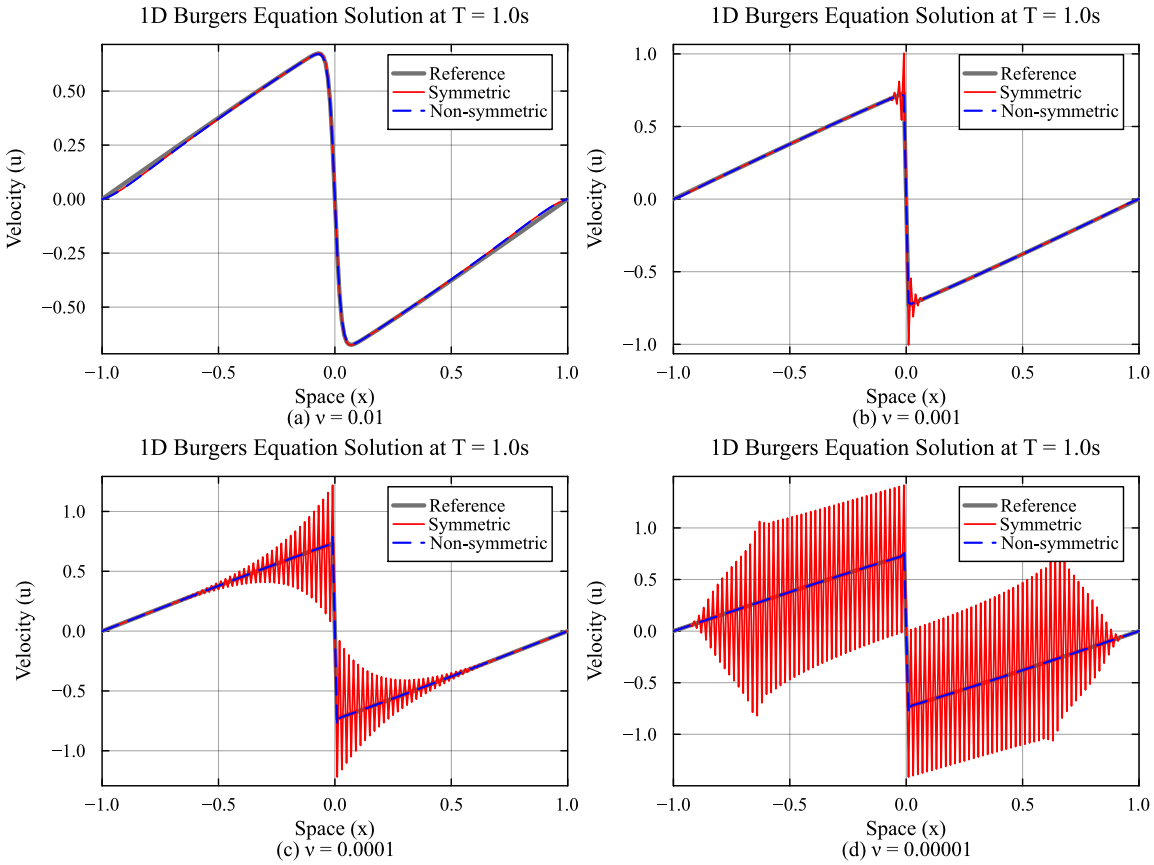


Figure 2: Posterior velocity profiles near the shock ($x \in [-0.1, 0.1]$) at $t = 1$ for decreasing viscosity ν.

4.3. Uncertainty Quantification

Beyond the point estimate, we assess whether the posterior covariance faithfully reflects the solution error. We use three measures, reporting coverage and the interval score at the

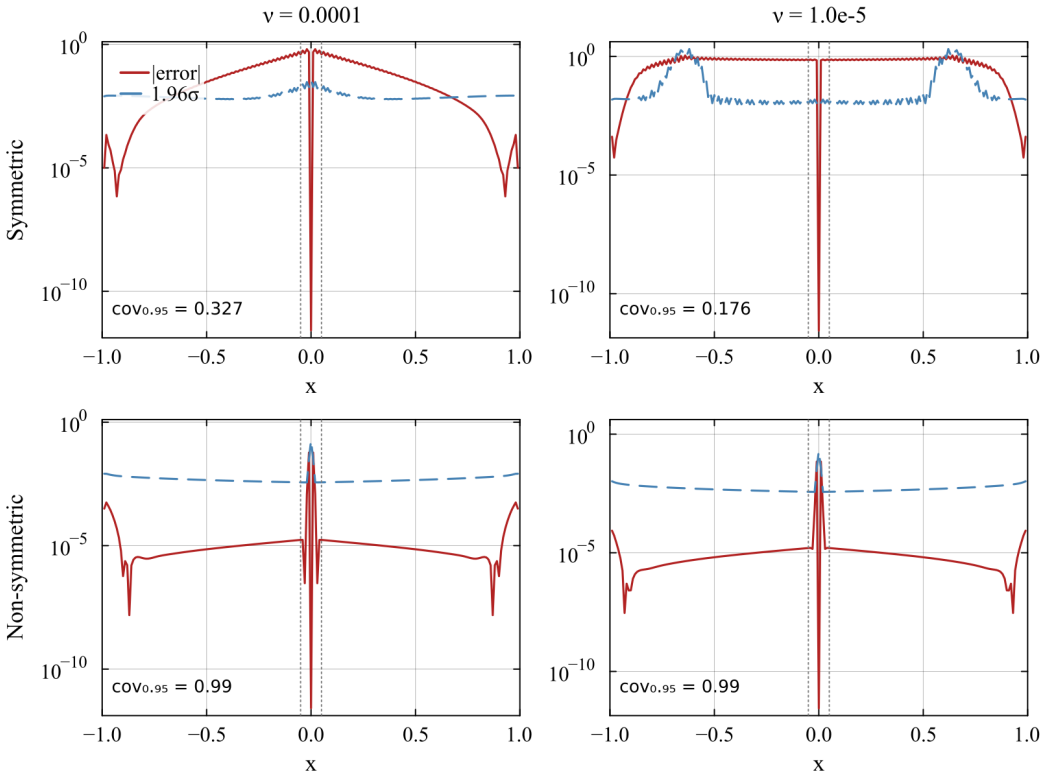


Figure 3: Pointwise error $|u - u^*|$ and 95% credible bound ($\pm 1.96\sigma$) at $t = 1$, log scale: symmetric (top) vs. non-symmetric (bottom), $\nu = 10^{-4}$ (left) and 10^{-5} (right).

same 95% level ($\alpha = 0.05$): the **empirical coverage** $\mathrm{cov}_{0.95}$, the fraction of nodes with the reference inside the 95% credible interval ($\pm 1.96\sigma$); the **ratio** $\bar{\sigma}/\mathrm{MAE}$, which equals $\sqrt{\pi/2} \approx 1.253$ for a correctly calibrated zero-mean Gaussian posterior; and the **interval score** (IS), which penalizes overconfident narrow intervals rather than rewarding them.

At the most demanding configuration ($\nu = 10^{-5}$, $\lambda_{\mathrm{pde}} = 10^6$, $\tau_p = 1$), the symmetric operator covers only 18% of interior nodes against 99% for the non-symmetric one, and is badly overconfident ($\bar{\sigma}/\mathrm{MAE} = 0.16$ versus the calibrated 1.253; Table 2). The mechanism follows from Section 4.2: the oscillatory solution has grid-scale gradients up to 2.57× the reference, which inflate the Gauss–Newton Jacobian; the linearization reads this as an abundance of information and underestimates the posterior variance exactly where the estimate is worst. Figure 3 makes this visible: for the symmetric operator the pointwise error escapes its own 95% credible band—near the shock at $\nu = 10^{-4}$ and across much of the domain at $\nu = 10^{-5}$—so that coverage drops from 0.33 to 0.18, whereas the non-symmetric operator's band encloses the error almost everywhere at both viscosities.

This behavior is robust across hyperparameters (Table 2). As the constraint precision λ_{pde} increases from 10^3 to 10^6 (Table 2(a)), the two operators respond oppositely. The non-symmetric posterior contracts cleanly as $\sigma \propto \lambda_{pde}^{-1/2}$, and its interval score improves by more than an order of magnitude. The symmetric operator's error, by contrast, is structural and independent of λ_{pde}, so its intervals contract at the same rate while covering ever less of the

Table 2: Posterior calibration under hyperparameter sweeps ($\nu = 10^{-5}$).

(a) Varying λ_{pde} ($\tau_p = 1$)

	Symmetric		Non-symmetric	
λ_{pde}	$\text{cov}_{0.95}$	IS	$\text{cov}_{0.95}$	IS
10^3	0.477	9.640	1.000	0.404
10^4	0.367	14.84	1.000	0.129
10^5	0.266	17.59	1.000	0.041
10^6	0.176	19.38	0.990	0.026

(b) Varying τ_p ($\lambda_{pde} = 10^6$)

	Symmetric		Non-symmetric	
τ_p	$\text{cov}_{0.95}$	$\bar{\sigma}$/MAE	$\text{cov}_{0.95}$	$\bar{\sigma}$/MAE
1	0.176	0.16	0.990	4.59
0.1	0.075	0.06	0.990	4.57
0.05	0.075	0.04	0.990	4.26
0.02	0.075	0.03	0.990	2.88
0.01	0.065	0.02	0.990	1.31

reference: coverage falls monotonically from 0.48 to 0.18 and the interval score deteriorates monotonically from 9.64 to 19.38.

The same pattern holds against the prior strength τ_p (Table 2(b)). As τ_p tightens from 1 to 10^{-2}, the non-symmetric coverage stays at 0.99 while $\bar{\sigma}$/MAE decreases from 4.59 towards the calibrated value 1.253, so the posterior scale is tunable without sacrificing coverage. The symmetric operator moves the other way: its coverage falls to 0.065 and it grows steadily more overconfident.

We therefore describe the default non-symmetric posterior as conservative rather than exactly calibrated: its intervals are wider than the error by roughly a factor of four, and although exact scale calibration is reachable by tightening the prior, it comes partly at the cost of the point estimate in the smooth region. The central finding is not that either posterior is perfectly calibrated, but that the two operators miscalibrate in opposite directions—the non-symmetric conservatively, the symmetric dangerously so where the error is largest—and that this contrast is robust across the tested hyperparameters.

5. Conclusion

We proposed a non-symmetric information operator for probabilistic inference in GMRFs that addresses the spurious oscillations standard symmetric formulations produce in advection-dominated regimes. By introducing a directional preference into the likelihood, the approach aligns the dependency structure of the posterior graph with the directed information flow of the underlying dynamics.

We showed that although the information operator is non-symmetric, the posterior precision matrix remains symmetric and positive definite, preserving the sparse Markov structure and ensuring compatibility with standard GMRF inference.

Experiments on the viscous Burgers' equation across four viscosity regimes showed that the method suppresses spurious oscillations and matches the finite-element reference. Moreover, its posterior uncertainty remains conservative across the tested hyperparameters, whereas the symmetric formulation is overconfident.

Extending the construction to higher-dimensional domains coupled PDE systems, and tightening the calibration, are the natural next steps.

Acknowledgments

This work was partially supported by the Ministry of Science, Innovation and Universities, Spain under Project AEI/10.13039/501100011033-PID2022-139977NB-I00, Project PLEC2023-010252/MIG-20232016 and also by the Autonomous Region of Madrid under Project ELLIS Unit Madrid and TEC-2024/COM-89. The authors gratefully acknowledge funding support from the China Scholarship Council (CSC) under the CSC-UPM Scholarship Project (Grant No.202506290101).

References

Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447:110668, 2021.

Jon Cockayne, Chris Oates, Tim Sullivan, and Mark Girolami. Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789, 2019.

Ramon Codina. Stabilized finite element approximation of transient incompressible flows using orthogonal subscales. *Computer Methods in Applied Mechanics and Engineering*, 191(39):4295–4321, 2002.

Oliver Hamelijnck, Arno Solin, and Theodoros Damoulas. Physics-informed variational state-space Gaussian processes. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2024.

Philipp Hennig, Michael A. Osborne, and Hans P. Kersting. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press, 2022.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Dinesh Kumar, Farid Ahmed, Shoaib Usman, Ayodeji Alajo, and Syed Alam. Recent advances in uncertainty quantification methods for engineering problems. *arXiv preprint arXiv:2211.03012*, 2022.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73 (4):423–498, 2011.

Finn Lindgren, Haakon Bakka, David Bolin, Elias Krainski, and Håvard Rue. A diffusion-based spatiotemporal extension of Gaussian Matérn fields. *Statistics and Operations Research Transactions*, 48(1):3–66, 2024.

Nora Lüthen, Stefano Marelli, and Bruno Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification*, 9 (2):593–649, 2021.

Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*, 6:1256–1269, 2024.

Sixtine Michel, Davide Torlo, Mario Ricchiuto, and Rémi Abgrall. Spectral analysis of continuous FEM for hyperbolic PDEs: Influence of approximation, stabilization, and time-stepping. *Journal of Scientific Computing*, 89(31), 2021.

Marvin Pförtner, Ingo Steinwart, Philipp Hennig, and Jonathan Wenger. Physics-informed Gaussian process regression generalizes linear PDE solvers. *arXiv preprint arXiv:2212.12474*, 2022.

Farzin Shakib, Thomas J. R. Hughes, and Zdeněk Johan. A new finite element formulation for computational fluid dynamics: X. the compressible Euler and Navier–Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 89(1–3):141–219, 1991.

Yan Shi, Pengfei Wei, Ke Feng, De-Cheng Feng, and Michael Beer. A survey on machine learning approaches for uncertainty quantification of engineering systems. *Machine Learning for Computational Science and Engineering*, 1(1):11, 2025.

Frank Shih, Zhenghao Jiang, and Faming Liang. Uncertainty quantification for physics-informed neural networks with extended fiducial inference. In *Advances in Neural Information Processing Systems*, pages 75361–75397. Curran Associates, Inc., 2025.

Tayfun E. Tezduyar and Yasuo Osawa. Finite element stabilization parameters computed from element matrices and vectors. *Computer Methods in Applied Mechanics and Engineering*, 190(3–4):411–430, 2000.

Tim Weiland, Marvin Pförtner, and Philipp Hennig. Flexible and efficient probabilistic PDE solvers through Gaussian Markov random fields. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 2746–2754. PMLR, 2025.

Dongbin Xiu and George Em Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

Liu Yang, Xuhui Meng, and George Em Karniadakis. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

Appendix A. Symmetry and Positive Definiteness of the Posterior Precision Matrix

Proposition 1 *Let $\tilde{\boldsymbol{A}} \in \mathbb{R}^{M\times N}$ be an arbitrary real matrix, $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ a symmetric positive definite matrix, and $\boldsymbol{Q}_\epsilon \in \mathbb{R}^{M\times M}$ a symmetric positive definite matrix. Then the posterior precision matrix $\boldsymbol{Q}_{\text{post}} = \boldsymbol{Q} + \tilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon \tilde{\boldsymbol{A}}$ is symmetric and positive definite.*

Proof *Symmetry.* Taking the transpose and using $\boldsymbol{Q}^\top = \boldsymbol{Q}$ and $\boldsymbol{Q}_\epsilon^\top = \boldsymbol{Q}_\epsilon$,

$$\boldsymbol{Q}_{\text{post}}^\top = \boldsymbol{Q}^\top + \left(\tilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon \tilde{\boldsymbol{A}}\right)^\top = \boldsymbol{Q} + \tilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon \tilde{\boldsymbol{A}} = \boldsymbol{Q}_{\text{post}}. \tag{22}$$

Positive definiteness. For any non-zero $\boldsymbol{v} \in \mathbb{R}^N$,

$$\boldsymbol{v}^\top \boldsymbol{Q}_{\text{post}}\, \boldsymbol{v} = \underbrace{\boldsymbol{v}^\top \boldsymbol{Q}\, \boldsymbol{v}}_{>0} + \underbrace{(\tilde{\boldsymbol{A}}\boldsymbol{v})^\top \boldsymbol{Q}_\epsilon (\tilde{\boldsymbol{A}}\boldsymbol{v})}_{\geq 0} > 0. \tag{23}$$

The first term is strictly positive because $\boldsymbol{Q} \succ 0$ and $\boldsymbol{v} \neq \boldsymbol{0}$. The second term is a quadratic form in $\boldsymbol{Q}_\epsilon \succ 0$ and is therefore non-negative for every $\boldsymbol{v}$; it may vanish when $\tilde{\boldsymbol{A}}$ has a non-trivial kernel, but this does not affect the strict inequality, which is already guaranteed by the first term. Hence $\boldsymbol{Q}_{\text{post}} \succ 0$. ■