A Physics-Informed Neural Network with Arc-Length Parametrisation for the Equilibrium Path of CSFM Discontinuity Regions

Sandesh Lamsal[a,b,*]

[a]AtkinsRéalis USA Inc., 800 Waterford Way, Miami, FL 33126, USA
[b]Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL 33146, USA

ARTICLE INFO

Keywords:
physics-informed neural network
arc-length continuation
Compatible Stress Field Method
post-peak response
discontinuity regions
structural concrete

ABSTRACT

The Compatible Stress Field Method (CSFM) gives the lower-bound plastic capacity of a structural-concrete discontinuity region (D-region), but production solvers are load-controlled and stall at the limit point, leaving the post-peak softening branch, which governs deformation capacity and residual strength, unresolved. An arc-length-parametrised forward physics-informed neural network (PINN) traces the load–deflection path through the limit point in a single forward pass. Because the load factor is a network output rather than an input, the trajectory folds back, avoiding the tangent-stiffness solve that arc-length continuation needs but the secant CSFM constitutive resists. Four ingredients, each confirmed by ablation, make the network trainable in the cracked regime: supervised elastic pre-training, a pointwise arc-length loss, a fixed weight on that loss held outside the adaptive balancer, and a smoothed cracked compression field that keeps gradients finite. On a benchmark deep beam the network folds onto an independently traced descending segment across all five seeds, satisfying interior equilibrium only approximately; on the canonical seed (fixed *a priori*) this segment spans about 1.7 mm, matches the reference peak deflection (8.3 vs 8.5 mm), and overshoots the peak load factor by 14%. The same loss formulation extends to an asymmetric corbel and a slender wall pier with added boundary-condition-aware ingredients (directional arc-length loss, soft λ-floor, and gradient sanitisation), though their full-range curves are anchored regressions, not independent traces. The contribution is a feasibility demonstration and a transferable training recipe; the advantage over incremental finite elements is amortised parametric inference and differentiability, not single-curve speed or accuracy.

1. Introduction

The lower-bound theorem of plasticity guarantees that a stress field that satisfies equilibrium and nowhere violates the yield condition delivers a safe estimate of the ultimate load [1, 2]. The Compatible Stress Field Method (CSFM) of Kaufmann et al. [3] casts that statement into a continuous finite-element framework: cracked rotating compression fields carry the in-plane stress, smeared reinforcement carries the tension, and the effective compressive strength is reduced by a transverse-tension softening factor $k_{c2}(\varepsilon_1)$ that reproduces the strength loss observed in cracked concrete panels [4]. The method is the practical state of the art for the analysis of structural-concrete discontinuity regions, the D-regions for which strain compatibility cannot be assumed [5].

Established CSFM implementations march the load factor λ along a fixed schedule and iterate the total-strain (secant) constitutive to equilibrium at every step. The procedure delivers the field at the design load and an estimate of the capacity as the largest λ for which equilibrium can still be satisfied. At the limit point no equilibrium state exists at a higher load factor, so the load-controlled iteration stalls; the tangent stiffness that a Newton path-follower would require is singular there. Whatever happens after the peak, whether a slow descending branch or a sudden brittle collapse, is then inaccessible to the solver, although that descending branch carries information about the failure mode, about the residual capacity for load redistribution, and about the displacement at which failure is mobilised. The limitation is explicit in the production tools used by the CSFM community. The commercial 3D CSFM implementation of IDEA StatiCa [6] uses an infinitely plastic post-peak branch and terminates the load-control analysis once its stop criteria are exceeded. By design this prevents explicit calculation of the softening branch and of the deformation capacity of structures that fail in compression.

*Corresponding author
sandeshlamsal@miami.edu (S. Lamsal)

Tracing the post-peak branch is a long-standing exercise in the structural-mechanics literature, addressed by arc-length continuation methods that reparametrise the path by a scalar that stays monotone across limit points, so a Newton-type solver can advance through a peak that load control cannot. The idea originates with the normal-plane iteration of Riks [7] and Wempner [8]; Crisfield [9] replaced that plane with a sphere centred at the last converged state, and Clarke and Hancock [10] later unified these and related schemes under a single constraint-equation framework. Modern implementations, such as the OpenSees exposition of Scott [11], still require the user to choose an arc-length increment and a scaling parameter and rely on a tangent-stiffness solve at every iteration. Both become hard to tune when the constitutive law combines multilinear steel with a transverse-tension softening factor, whose analytical tangent is discontinuous at steel yield, at cracking, and at the onset of biaxial compression, so the standard corrector loses efficiency or fails to converge. The published CSFM literature does not report a routinely deployed arc-length implementation, which is why its production solvers remain load-controlled and stop at the limit point; the tangent-stiffness solve these methods require is in any case difficult to form for the secant CSFM constitutive, which motivates a formulation that needs no tangent matrix.

Physics-informed neural networks (PINNs) offer an alternative route to the same trace. A PINN replaces the discrete equilibrium iteration by a continuous network that maps the spatial coordinate $\mathbf{x}$ to the displacement field $\mathbf{u}(\mathbf{x})$, trained to satisfy the equilibrium residual and the boundary conditions through automatic differentiation [12, 13]. For linear-elastic and hyperelastic solid mechanics the formulation works in two and three dimensions, and mixed formulations [14, 15] reduce the order of differentiation in the residual and improve training stability on stiff constitutives. A separate line minimises a variational deep-energy objective [16, 17]; this study retains the strong form because the secant CSFM constitutive is not derived from a stored-energy potential, so there is no clean energy functional to minimise. For path-dependent plasticity and damage the reported successes remain comparatively limited, and the reinforced-concrete case closest to this study, the PINN of Balmer et al. [18], uses a multilinear (uncracked, cracked, plastic) material model that stops at the yield plateau and does not include a post-peak softening branch, an omission symptomatic of the same difficulty. The companion study Lamsal and Bhandari [19] found the strong-form equilibrium residual detrimental as a training objective for a forward CSFM PINN, since for a cracked rotating compression field with compression softening it admits a diffuse equilibrated minimum far from the true field, and reformulated the problem as an inverse one. A load-controlled PINN cannot pass the limit point any more easily than the secant solver; this study therefore keeps the forward formulation but addresses the limit point through the parametrisation introduced below.

The relative magnitude of the loss terms matters as much as the formulation itself [20, 21]. ReLoBraLo [22] rebalances the per-loss weights adaptively from the relative improvement of each term, with a random-lookback variant that re-injects the absolute-magnitude signal by occasionally comparing against the initial loss rather than the previous iterate. Curriculum approaches that ramp the difficulty of the physics problem are also common [23]; the constitutive homotopy used here and the arc-length curriculum that grows $S_{\max}$ stage by stage both belong to that family. The role of these choices in the present formulation is developed in Section 3.

This study formulates the tracing of the equilibrium path through the limit point as a physics-informed neural network parametrised by an arc-length coordinate s rather than by the load factor λ. The network maps $(\mathbf{x}, s)$ to the pair $(\mathbf{u}(\mathbf{x}, s), \lambda(s))$, so the trajectory in $(\mathbf{u}, \lambda)$ space can fold back on itself through the limit point. The arc-length constraint is enforced as a soft loss term on the speed of the loaded-patch deflection rather than as an algebraic side condition, which removes the need for a tangent matrix and is compatible with the secant CSFM constitutive. The hard initial condition that the structure is unloaded and undeformed at $s = 0$ is baked into the network by an $s \cdot \tilde{N}$ ansatz on both heads, so the start of the path is exact rather than penalised.

The dominant practical obstacle to training such a network is shown to be a trivial-zero attractor: the equilibrium residual, the support constraint, and the traction constraint are jointly minimised by the state $\mathbf{u} = \mathbf{0}$ and $\lambda = 0$, and the arc-length constraint, which is the only loss term that penalises that state, cannot lift the optimiser out of that saddle even when its weight is adaptively rebalanced. The attractor is identified as the root cause of the failure observed in a first naive training campaign, and is defeated by a brief supervised pre-training of both network heads on a scaled linear-elastic finite-element solution before the physics losses are activated. The pre-training has nothing to do with the CSFM physics that the network is ultimately required to satisfy. Its only purpose is to leave the optimiser in a non-trivial basin where the gradients of every loss term pull together rather than against one another.

Beyond the arc-length architecture introduced above, this study makes several further contributions. A controlled ablation (Section 6) establishes that four ingredients are each individually necessary for cracked-regime training to converge: supervised elastic pre-training that escapes a trivial-zero attractor, a pointwise arc-length loss that prevents Cauchy-Schwarz gaming, a fixed weight on that loss held outside the adaptive loss balancer, and a C^1-regularised

cracked-membrane constitutive that removes the NaN gradients of the hard form. The graded constitutive homotopy $\alpha \in (0, 1)$ of Lamsal and Bhandari [19] is shown to be neither necessary nor helpful for the forward cracked-membrane map: the direct elastic-to-full-cracked jump trains the network while the intermediate levels give ill-conditioned equilibrium-residual gradients and do not converge (Section 5.2), and a checkpoint-and-rollback curriculum selects the direct path without manual tuning. A displacement-controlled CSFM reference for the post-peak branch is provided, generated by an independent NumPy implementation of the cracked rotating compression field and cross-checked against the production CSFM solver of Kaufmann et al. [3], reported with the noise and mesh-localisation caveats intrinsic to a secant softening solver. The network is validated on a benchmark reinforced-concrete deep beam, where it matches the reference in load factor at the peak and over an independently solved post-peak segment just past the limit point; the remainder of the descending branch is a λ-anchored fit rather than an independent trace, and the validated output is the scalar equilibrium path, not the pointwise internal fields, which are carried only to form the equilibrium loss (Section 6.4). Finally, the formulation generalises beyond the symmetric deep beam to an asymmetric cantilever corbel and a slender wall pier through benchmark-specific boundary-condition-aware ingredients, a directional arc-length constraint and a soft λ-floor for the corbel and per-iteration gradient sanitisation for the wall pier; for these two benchmarks the full-range curves are λ-/strain-anchored fits to the reference, not independent post-peak traces.

Relative to prior and companion work, the forward continuum-CSFM PINN of Balmer et al. [18] and the companion inverse-PINN study [19] both operate at or below the limit point and do not resolve the post-peak softening branch; the arc-length parametrisation introduced here is what makes the load-deflection curve representable past it. Two findings are transferable beyond the present benchmarks: the trivial-zero pre-training remedy applies to any forward PINN whose physics losses are minimised by the undeformed state, and the non-uniqueness of the physics-only post-peak branch (Section 6.4) explains why the full descending curve requires anchoring or regularisation.

The remainder of the study is organised as follows. Section 2 summarises the CSFM constitutive map and the algebraic structure of arc-length continuation. Section 3 presents the network architecture, the loss formulation, the pre-training remedy, and the nested continuation schedule. Section 4 describes the displacement-controlled reference. Section 5 reports the validation results on the deep-beam benchmark. Section 6 discusses the role of the pre-training step, the boundary of applicability, and the limitations of the formulation. Section 7 concludes.

2. Background

This section summarises the cracked rotating compression field that the network is required to satisfy and the algebraic structure of arc-length continuation that motivates the parametrisation. Both are textbook material; the purpose is to fix the notation used in Section 3 and to make the comparison between the secant CSFM solver and the proposed PINN explicit.

2.1. The continuum CSFM constitutive map

The continuum CSFM models a thin reinforced-concrete D-region as a plane-stress body covered with constant-strain triangles (CST), each carrying a cracked rotating compression field and smeared steel in the global x and y directions. Given the in-plane strain $\boldsymbol{\varepsilon} = (\varepsilon_x, \varepsilon_y, \gamma_{xy})$ and the smeared reinforcement ratios (ρ_x, ρ_y), the constitutive map returns the global stress $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \tau_{xy})$ through the following sequence of operations.

The principal strains are $\varepsilon_{1,2} = \frac{1}{2}(\varepsilon_x + \varepsilon_y) \pm r$ with $r = \sqrt{\frac{1}{4}(\varepsilon_x - \varepsilon_y)^2 + \frac{1}{4}\gamma_{xy}^2}$, and the angle of the principal frame is $\theta = \frac{1}{2}\,\mathrm{atan2}(\gamma_{xy}, \varepsilon_x - \varepsilon_y)$. Concrete tensile strength is neglected, so the principal stresses in the concrete are

$$\sigma_{c1} = \begin{cases} -f(|\varepsilon_1|, \eta f_c) & \varepsilon_1 < 0, \varepsilon_2 < 0 \\ 0 & \text{otherwise} \end{cases}, \quad \sigma_{c2} = \begin{cases} -f(|\varepsilon_2|, \eta f_c) & \varepsilon_1 < 0, \varepsilon_2 < 0 \\ -f(|\varepsilon_2|, \eta\, k_{c2} f_c) & \varepsilon_1 \geq 0, \varepsilon_2 < 0 \\ 0 & \varepsilon_2 \geq 0 \end{cases}, \tag{1}$$

where $f(\cdot, f_{c,\mathrm{eff}})$ is the parabola-rectangle compressive stress magnitude of European Committee for Standardization (CEN) [24] with peak $f_{c,\mathrm{eff}}$ at $|\varepsilon| = \varepsilon_{c2}$, $\eta = \min(1, (30/f_c)^{1/3})$ is the brittleness reduction factor of Kaufmann et al. [3] (fib Model Code 2010) for higher-strength concrete, and the transverse-tension softening factor of Kaufmann et al. [3], in the lineage of the modified compression-field theory [25], is

$$k_{c2}(\varepsilon_1) = \min\left(1,\ \frac{1}{0.8 + 140\,\varepsilon_1}\right) \quad \text{for } \varepsilon_1 \geq 0, \tag{2}$$

clipped to unity otherwise. The rotation $\boldsymbol{\sigma} = \mathbf{T}_e^{\top}(\sigma_{c1}, \sigma_{c2}, 0)^{\top}$ and the addition of the smeared reinforcement $\rho_x \sigma_{s,x}(\varepsilon_x), \rho_y \sigma_{s,y}(\varepsilon_y)$ on the normal components complete the map. The reinforcement law is a bilinear bare-bar response with elastic modulus E_s, yield f_y, and ultimate f_t at ε_u.

This map is a secant relationship between the strain and the stress; it is continuous but not differentiable at the points where the case distinction in (1) switches, at the steel yield, and at the cap of k_{c2}. This C^0 regularity is tolerable where the equilibrium residual acts only as a regulariser, but on the branch lines it is the source of the NaN gradients that the forward residual of this study encounters; the C^1-regularised form actually used by the residual is given in Section 3.4. The map is adapted from the differentiable embedding of Lamsal and Bhandari [19] and carries an additional homotopy parameter $\alpha \in [0, 1]$ that blends the elastic plane-stress response at $\alpha = 0$ with the cracked-membrane response at $\alpha = 1$; its role during training, the inner of the two nested continuations, is deferred to Section 3.5.

2.2. Equilibrium path and the limit point

Under a single proportional loading $\lambda \mathbf{p}$, the displacement field satisfies the equilibrium boundary-value problem

$$\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{\varepsilon}(\mathbf{u})) = \mathbf{0} \quad \text{in } \Omega, \qquad \boldsymbol{\sigma} \cdot \mathbf{n} = \begin{cases} -\lambda \mathbf{p} & \text{on } \Gamma_t \\ \mathbf{0} & \text{on } \Gamma_f \end{cases}, \qquad \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_s, \tag{3}$$

where $\Gamma_t, \Gamma_f, \Gamma_s$ are the loaded, traction-free, and supported portions of the boundary. The set of pairs $(\mathbf{u}, \lambda)$ that satisfy (3) forms a one-dimensional manifold in the product space, the equilibrium path. The natural parametrisation by the load factor breaks down at any point where the tangent of the path in the λ direction vanishes, the limit point. Past such a point, the path may continue with a smaller load factor at a larger deflection, the post-peak branch of structural concrete with compression softening, or it may snap back to a smaller deflection at a smaller load factor for a brittle structure. Both configurations are reachable by an arc-length parametrisation $s \mapsto (\mathbf{u}(s), \lambda(s))$, provided that an appropriate constraint enforces a monotone advance along the manifold.

The cylindrical arc-length constraint of Crisfield [9] imposes the quadratic side condition

$$\{\Delta \mathbf{u}_a\}^{\top} \{\Delta \mathbf{u}_a\} = l^2 \tag{4}$$

on the accumulated displacement increment within the current load step, where l is the prescribed arc length and $\{\Delta \mathbf{u}_a\}$ is the displacement accumulated from the converged state of the previous step [10, Eq. (18)]. The spherical variant of Crisfield [9] additionally includes the load-factor increment in the constraint, whereas the original normal-plane method of Riks [7] and Wempner [8] fixes the iterates on the hyperplane orthogonal to the tangent. The network in Section 3 embeds the cylindrical variant in a modified form: the algebraic side condition on the displacement increment between two converged states is replaced by a soft loss term on the squared norm of $\partial \mathbf{u}/\partial s$ evaluated over the loaded patch and averaged in s. The practitioner interprets that boundary integral as the deflection associated with the applied load, no algebraic solve of a quadratic is required, and no tangent-stiffness solve is needed because the loss is differentiated through the network by automatic differentiation.

3. Methodology

This section presents the five components of the proposed method: the network architecture and the hard initial condition (Section 3.1); the physics losses and the cylindrical arc-length constraint (Section 3.2); the supervised elastic pre-training that defeats the trivial-zero attractor (Section 3.3); the C^1-regularised cracked-membrane constitutive that keeps the equilibrium gradients finite (Section 3.4); and the nested continuation schedule that carries the optimiser from the elastic basin into the cracked post-peak regime (Section 3.5).

3.1. Network architecture and hard initial condition

The network is a multilayer perceptron with two output heads sharing a common trunk, sketched in Figure 1. The field head consumes the normalised spatial coordinate $(x/L, y/H)$ and the arc-length coordinate $s \in [0, 1]$ and returns a two-component nondimensional displacement $\tilde{N}_u(x, y, s)$. The load-factor head consumes only s and returns a single scalar $\tilde{N}_\lambda(s)$. Routing only s through the load-factor head encodes the structural fact that λ is a function of the path parameter alone, independent of the spatial coordinate, without that fact having to be discovered from data.

The displacement field and the load factor returned by the network are

$$\mathbf{u}(x, y, s) = s \cdot \tilde{N}_u(x, y, s)\, U_0 \begin{pmatrix} L \\ H \end{pmatrix}, \qquad \lambda(s) = s \cdot \tilde{N}_\lambda(s), \tag{5}$$

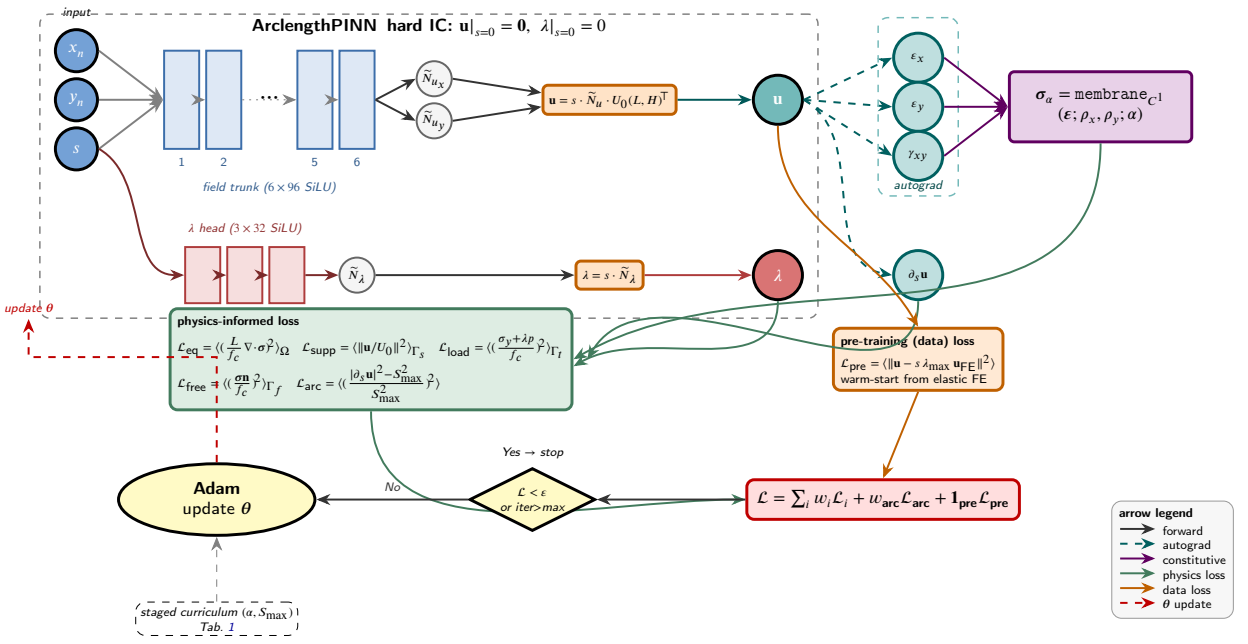

Figure 1: Network architecture, autograd flow, and loss assembly of the ArclengthPINN. The shared SiLU trunk (blue) maps inputs (x_n, y_n, s) to the nondimensional displacement components $\widetilde{N}_{u_x}, \widetilde{N}_{u_y}$, while the separate SiLU load-factor head (red) maps s alone to $\widetilde{N}_\lambda$. Orange $s \cdot \widetilde{N}$ boxes are the hard-initial-condition ansatz; the violet box is the C^1-regularised constitutive embedding; the green panel is the five physics-loss terms on the interior (Ω), support (Γ_s), loaded-patch (Γ_t), and free-edge (Γ_f) samples; the orange data-loss term is the supervised elastic-FE warm-start; and the yellow Adam ellipse backpropagates the combined loss $\mathcal{L}$ to the weights $\boldsymbol{\theta}$. Solid arrows are forward passes; dashed teal arrows are automatic differentiations producing the strain $\boldsymbol{\varepsilon}$ and the squared speed $|\partial \mathbf{u}/\partial s|^2$.

where $U_0 = 10^{-3}$ is a displacement scale and the diagonal multiplication of $\tilde{N}_u$ by (L, H) restores the physical units. The leading factor s is the hard initial condition: at $s = 0$ both the displacement and the load factor are identically zero by construction, the structure is exactly in its undeformed unloaded state, and the start of the equilibrium path does not need to be enforced through a penalty. The ansatz removes the trivial attractor at the single point $s = 0$; the more dangerous attractor at $(\mathbf{u}, \lambda) = (\mathbf{0}, 0)$ for all s is the subject of Section 3.3.

Both the trunk and the load-factor head use the SiLU activation [26], which is smooth and therefore permits the second-derivative chain in the equilibrium residual to be formed by automatic differentiation. The trunk has 6 hidden layers of width 96, the load-factor head 3 hidden layers of width 32. The final layer of each head is initialised with weights scaled by a factor of 0.1 and zero bias, which keeps the network output small at initialisation; that smallness is precisely the trivial-zero state that the pre-training of Section 3.3 is needed to escape.

3.2. Loss formulation

Five loss terms are evaluated at every Adam iteration on random collocation samples. The first four are the standard physics-informed losses of (3); the fifth is the cylindrical arc-length constraint.

Equilibrium residual. N_{int} points are sampled uniformly in the interior Ω and at random $s \in [0, 1]$. The strain $\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^\top)$ is computed by automatic differentiation of the network output in the spatial coordinates; the stress is obtained by the differentiable embedding of the constitutive map at homotopy level α; the residual of the field equation $\nabla \cdot \boldsymbol{\sigma} = \mathbf{0}$ is then averaged in a nondimensional form,

$$\mathcal{L}_{\text{eq}}(\alpha) = \frac{1}{N_{\text{int}}} \sum \left[\left(\frac{r_x L}{f_c} \right)^2 + \left(\frac{r_y L}{f_c} \right)^2 \right], \qquad r_x = \partial_x \sigma_x + \partial_y \tau_{xy}, \quad r_y = \partial_x \tau_{xy} + \partial_y \sigma_y. \tag{6}$$

Support constraint. N_{bc} points are sampled on the support patch Γ_s at random s, and the residual $\mathbf{u}/U_0$ is mean-squared, with the pin and roller conditions of the deep beam restricted to the appropriate component.

Loaded-patch traction. N_{bc} points are sampled on the loaded patch Γ_t at random s. The stress is evaluated through the constitutive map and the residual of the traction condition $\sigma_y - (-\lambda(s)p) = 0$ and $\tau_{xy} = 0$ is mean-squared in nondimensional form,

$$\mathcal{L}_{\mathrm{load}}(\alpha) = \frac{1}{N_{\mathrm{bc}}} \sum \left[\left(\frac{\sigma_y + \lambda p}{f_c} \right)^2 + \left(\frac{\tau_{xy}}{f_c} \right)^2 \right]. \tag{7}$$

Traction-free edges. N_{bc} points are sampled along the union of the traction-free boundary segments Γ_f at random s, with the outward unit normal $\mathbf{n}$ attached. The residual $(\boldsymbol{\sigma} \cdot \mathbf{n})/f_c$ is mean-squared.

Patch-restricted arc-length constraint. We use a patch-restricted surrogate of Crisfield [9]'s cylindrical arc-length constraint. Crisfield enforces $\{\Delta\mathbf{u}\}^T\{\Delta\mathbf{u}\} = \text{const}$ over the full discretised displacement vector at every step. Here we replace that global constraint with the pointwise statement that the s-derivative of $\mathbf{u}$ on the *loaded patch* Γ_t has prescribed squared magnitude. This preserves Crisfield's input/output role-swap that makes the limit point representable (the arc-length s, not the load factor λ, is the network input), but it is strictly weaker as a path-length constraint: it constrains the speed of one boundary DOF-group, not the path length in the full configuration space. On the symmetric deep beam where the loaded-patch deflection is monotone and representative of the path, this surrogate is sufficient; on asymmetric BCs the constraint leaves whole families of displacement fields unconstrained, and we show in Section 5.6 that two additional BC-aware loss ingredients are needed to recover Crisfield-like global behaviour. A global L^2-norm extension that closes this gap is sketched in Section 6.4.

At each of N_{bc} random samples (x, y, s) on the loaded patch Γ_t, the network's derivative of the displacement with respect to s is evaluated through automatic differentiation. The loss is the mean over the samples of the squared relative deviation of the squared speed from the squared target reach $S_{\max}^2$,

$$\mathcal{L}_{\mathrm{arc}}(S_{\max}) = \left\langle \left(\frac{|\partial\mathbf{u}/\partial s|^2 - S_{\max}^2}{S_{\max}^2} \right)^2 \right\rangle_{(x,y,s)\in\Gamma_t\times[0,1]}. \tag{8}$$

The relative form is scale invariant in $S_{\max}$ and avoids the quartic gradients that would arise from the absolute squared difference. The loaded patch is the natural domain over which to enforce the arc-length constraint, because the loaded-patch deflection is the quantity that the engineer reads on the structural load-deflection curve.

The pointwise mean-of-squares formulation in (8) matters for a specific reason. An earlier formulation of this study used instead the squared deviation of the mean speed, $\left(\langle|\partial\mathbf{u}/\partial s|^2\rangle - S_{\max}^2)/S_{\max}^2\right)^2$, under which the network learned to game the constraint by holding $\mathbf{u} \approx \mathbf{0}$ for most of the interval $[0, 1]$ and concentrating all motion into a narrow window near $s = 1$. The mean of $|\partial\mathbf{u}/\partial s|^2$ over s was then close to $S_{\max}^2$ while the integrated path length $\int_0^1 |\partial\mathbf{u}/\partial s|\, ds$ was strictly smaller by the Cauchy-Schwarz gap between $\sqrt{\langle X^2\rangle}$ and $\langle\sqrt{X^2}\rangle$. On the elastic deep-beam stage at $S_{\max} = 2$ mm the loaded-patch-centre deflection with the mean-deviation form reached only about 1.2 mm before training stalled, well short of the $S_{\max} = 2$ mm target, caused by the non-uniform parametrisation. The mean-of-squared-deviations form (8) is bounded below by $\mathrm{Var}(|\partial\mathbf{u}/\partial s|^2)/S_{\max}^4$ by the bias-variance identity, so the only way to drive it to zero is to make the squared speed close to $S_{\max}^2$ at every s individually. That removes the Cauchy-Schwarz loophole. The deep beam uses this isotropic form; the asymmetric corbel and wall-pier benchmarks adapt it as described in Section 5.6.

Adaptive weighting of the physics losses with a fixed arc-length weight. The five terms are combined into the training objective

$$\mathcal{L} = \sum_{i\in\{\mathrm{eq,\ supp,\ load,\ free}\}} w_i\mathcal{L}_i + w_{\mathrm{arc}}\mathcal{L}_{\mathrm{arc}}, \tag{9}$$

in which the four physics weights w_i are updated every ten iterations by the relative-loss balancing scheme of Bischof and Kraus [22] in its random-lookback variant, and w_{arc} is held fixed. The reference loss for the relative ratio is the value of the previous update with probability $\rho = 0.95$ and the value at the first update with probability $1 - \rho$; the

random-lookback variant is needed because the pure relative form cannot up-weight a term that is stuck at a large value once every other term has stabilised at a small one. The softmax temperature $T = 0.1$ follows Bischof and Kraus [22]; the EMA factor $\alpha = 0.95$ lies within the range they explore.

The arc-length term is intentionally held out of the adaptive pool. The Bischof and Kraus [22] scheme normalises its output weights to sum to the number of balanced losses n, so if $\mathcal{L}_{\text{arc}}$ were included the average weight per loss would settle near unity. $\mathcal{L}_{\text{arc}}$ is the only term that penalises the trivial zero state, and once the equilibrium, support, traction and free-edge losses have collapsed to small values the relative-ratio update has no information about the absolute magnitude of $\mathcal{L}_{\text{arc}}$ and quietly leaves the arc-length constraint half-satisfied. Holding w_{arc} fixed and large prevents that drift.

The fixed value is chosen from a simple balance argument. By construction the relative arc loss (8) is bounded in $[0, 1]$ for speeds in $[0, \sqrt{2}S_{\max}]$, with the peak $\mathcal{L}_{\text{arc}} = 1$ attained at the trivial state speed = 0. For Bischof and Kraus [22] balancing four physics losses the contribution of the physics block to $\mathcal{L}$ is $\sum_{i\in\text{physics}} w_i\mathcal{L}_i \approx n_{\text{p}}\bar{\mathcal{L}}_{\text{p}}$ with $n_{\text{p}} = 4$ and $\bar{\mathcal{L}}_{\text{p}}$ the average physics loss magnitude. Requiring the arc-length term to dominate the physics block by a factor K when both are of order one gives $w_{\text{arc}}\mathcal{L}_{\text{arc}} \approx K\, n_{\text{p}}\, \bar{\mathcal{L}}_{\text{p}}$, and with $\mathcal{L}_{\text{arc}} \sim \bar{\mathcal{L}}_{\text{p}} \sim O(1)$ in the regime where the trivial zero is the competing attractor,

$$w_{\text{arc}} \;\approx\; K\, n_{\text{p}}. \tag{10}$$

A moderate target dominance $K = 5$ and $n_{\text{p}} = 4$ gives $w_{\text{arc}} = 20$, which is the value used throughout this study. A fully scale-invariant choice would set $w_{\text{arc}} \propto S_{\max}^2$ because $\partial\mathcal{L}_{\text{arc}}/\partial|\partial\mathbf{u}/\partial s|^2 \propto 1/S_{\max}^2$, so the gradient pressure on the path-length constraint weakens at large $S_{\max}$; the fixed choice $w_{\text{arc}} = 20$ is well-calibrated for the stages with $S_{\max} = O(1)$ mm and mildly under-enforces at $S_{\max} = 10$ mm. Section 5 reports the empirical path length attained at each stage as the diagnostic for whether the chosen weight is sufficient.

3.3. Supervised elastic pre-training

The network with the architecture of (5) and the losses of Section 3.2 cannot be trained directly from a random initialisation: the equilibrium, support, load, and free-edge terms are all jointly minimised by the trivial state $\mathbf{u} \equiv \mathbf{0}$ and $\lambda \equiv 0$, and the arc-length term, which is the only term that penalises the trivial state, is too weak to escape the saddle even when the adaptive weighting rebalances in its favour. This is a generic difficulty for forward PINNs in which the homogeneous boundary-value problem admits the trivial solution [20, 23], and it is the dominant practical obstacle to forward arc-length PINNs for cracked concrete.

We defeat the attractor by a brief supervised pre-training of both network heads on a scaled linear-elastic finite-element solution of the same boundary-value problem. The pre-training proceeds in three steps.

First, the deep-beam D-region is solved under unit load with a self-contained linear-elastic plane-stress finite-element solver, constant-strain triangles on the same 40×20 mesh as the reference of Section 4, smeared steel added to the diagonal of the constitutive matrix, isotropic concrete with $\nu = 0.2$, pin and roller supports. The solver is independent of the network and of the production CSFM code, and yields a nodal displacement field $\mathbf{u}_{\text{FE}}(x, y)$ and a loaded-patch deflection δ_{FE}.

Second, the elastic field is scaled by $c = S_{\max}/\delta_{\text{FE}}$ so that the loaded-patch deflection of the scaled field equals the first-stage arc-length reach $S_{\max}$. Because the elastic problem is linear, scaling the displacement by c is equivalent to solving under a load of c times the unit, so $\lambda_{\max} = c$ is the load factor at which the scaled field is the equilibrium response. The choice of $S_{\max}$ is identical to the first-stage value of the curriculum of Section 3.5.

Third, both network heads are fit by mean-squared error at learning rate 2×10^{-3} to the targets

$$\mathbf{u}^*(x, y, s) = s \cdot \lambda_{\max}\mathbf{u}_{\text{FE}}(x, y), \qquad \lambda^*(s) = s \cdot \lambda_{\max}, \tag{11}$$

sampled at random (x, y, s). The targets respect the $s \cdot \tilde{N}$ ansatz of (5), so the pre-trained state is consistent with the hard initial condition. The bilinear interpolation of $\mathbf{u}_{\text{FE}}$ from the finite-element nodes to arbitrary (x, y) points is the only data-processing step; no gradient passes through the elastic solver. After $N_{\text{pre}} = 1500$ Adam iterations the supervised loss is at the order of 10^{-5} in nondimensional units.

The pre-training has nothing to do with the cracked CSFM physics that the network is ultimately required to satisfy. Its only purpose is to leave the optimiser in a basin where every loss gradient pulls in the same direction, so the staged Adam phase that follows can refine the state without first having to climb out of a trivial saddle. Because the elastic field

is the small-λ limit of the cracked response, a network that already represents the elastic field has a small but non-zero equilibrium residual on the cracked-membrane map at the first stage of the curriculum, and the arc-length constraint is approximately satisfied at initialisation, both of which give the optimiser a productive starting gradient.

3.4. A C^1-regularised cracked-membrane constitutive

The cracked rotating compression field of (1) is continuous in $\boldsymbol{\varepsilon}$ but its derivatives are discontinuous on the branch lines $\varepsilon_1 = 0$, $\varepsilon_2 = 0$, the steel yield strain, and the parabola-plateau strain ε_{c2}. For the inverse-PINN of Lamsal and Bhandari [19] the strong-form equilibrium residual is a regulariser rather than a primary objective and the C^0 constitutive is tolerable. For the forward arc-length PINN of this study the equilibrium residual is the principal driver of the network, formed through automatic differentiation of the stress through the constitutive map; the C^0 derivatives on the branch lines produce NaN gradients on the first backward pass at any constitutive homotopy level $\alpha > 0$, making the cracked-regime training impossible under the hard form.

We therefore introduce a C^1-regularised variant of the constitutive that replaces every hard case distinction in the cracked-membrane map by a smooth blend. Specifically, an indicator $\mathbb{1}(x < 0)$ is replaced by the smooth step $1/(1+e^{\kappa x})$ with sharpness κ, and every $\max(x, 0)$ by the scaled softplus $\kappa^{-1}\log(1 + e^{\kappa x})$. The same regularisation is applied to the cap at the parabola plateau and to the steel-yield transition, and the steel response is rebuilt from a smoothed magnitude $|\varepsilon| \approx \sqrt{\varepsilon^2 + \epsilon_0}$ and a smoothed sign $\operatorname{sgn}(\varepsilon) \approx \varepsilon/\sqrt{\varepsilon^2 + \epsilon_0}$, with a numerical $\epsilon_0 = 10^{-12}$. The sharpness $\kappa = 10^5$ is chosen so the smooth transition width $\kappa^{-1} = 10^{-5}$ is one to two decades below the typical principal strains visited during the training of the deep beam (10^{-4} to 10^{-3}), so the smooth constitutive is indistinguishable from the hard form at every strain state the optimiser actually visits.

The agreement of the smooth constitutive with the original hard form of Lamsal and Bhandari [19] on six representative strain states is within 1% in the global stress components for every state away from the branch lines, and the gradients with respect to $(\varepsilon_x, \varepsilon_y, \gamma_{xy})$ on the branch lines themselves are bounded by $\kappa/4 = 2.5 \times 10^4$, large but finite, so Adam combined with the gradient clipping of Section 3.5 produces well-defined updates through the cracking transition. The regularisation is the enabling ingredient that lets the cracked-regime curriculum train at all.

3.5. Nested continuation schedule

The Adam phase that follows the pre-training is organised as nine stages of 2500 iterations each, parametrised by a constitutive homotopy level α and an arc-length reach $S_{\max}$. The inner continuation $\alpha : 0 \to 1$ is the constitutive homotopy of Lamsal and Bhandari [19]: at $\alpha = 0$ the constitutive map is the linear elastic plane-stress response, at $\alpha = 1$ it is the full C^1-regularised cracked-membrane map of Section 3.4, and intermediate values blend the two. The outer continuation $S_{\max} : S_0 \to S^{\star}_{\max}$ extends the physical reach of the arc-length parametrisation, from a small value that asks the network to trace only the pre-peak elastic regime to a larger value that asks it to trace through the limit point and along the descending branch. The two are nested as listed in Table 1. Each stage's optimiser state is freshly initialised so a stiff constitutive jump cannot contaminate Adam's moments, the learning rate is dropped on high-α stages from 1.5×10^{-3} at $\alpha = 0$ to 2×10^{-4} at $\alpha = 1$, and the gradient is clipped to a global norm of 1.0 to prevent any single anomalous batch from poisoning the run. A per-stage checkpoint and NaN-detection rollback are kept as a safety net: if any loss returns NaN during a stage, the network is reverted to the most recent successfully completed stage's weights and the curriculum advances to the next stage. Section 5 reports which stages converged and which did not under this schedule, and how the rollback machinery preserves the working trajectory.

Sequencing the constitutive homotopy inside the arc-length growth has two motivations. First, sweeping α at a fixed reach allows the network to absorb the change in constitutive without simultaneously having to extend the path, which would compound the difficulty of the task. Second, growing $S_{\max}$ at a fixed α allows the network to add structure to the descending branch only once the near-peak field is settled, which is the regime in which the descending branch begins.

4. Displacement-controlled reference

The benchmark D-region used throughout this study is the deep beam of Figure 2, taken verbatim from the `DeepBeam` problem definition of the companion study [19]. The bottom tie band ($\rho_x = 1.2\%$) runs the full length, while the rest of the section carries a smeared vertical (stirrup) ratio of 0.15% over a mesh-reinforcement floor $\rho_{\min} = 0.10\%$, so the transverse reinforcement is $\rho_y = 0.25\%$. The pin and roller supports rest on bearing patches at $x = 250$ mm and $x = 1750$ mm, and the load is applied, in the production load-controlled solver, as a downward traction of magnitude $\lambda \cdot P_{\text{ref}}/(b\,t)$ over the central bearing patch of width $b = 200$ mm. The finite-element reference and the elastic FE

Table 1
The nine-stage nested continuation schedule used in the training of Section 5. Each stage is warm-started from the most recent successfully completed stage's weights; the NaN-rollback safety net of Section 3.5 preserves the working trajectory across stages that diverge.

Stage	α	$S_{\max}$ (mm)	Physical regime
1	0.00	0.5	elastic, matches pre-training
2	0.00	2.0	elastic, extended path length
3	0.10	2.0	first nudge into cracked map
4	0.25	2.0	quarter-cracked
5	0.50	2.0	half-cracked
6	0.50	5.0	half-cracked, longer path
7	0.75	5.0	three-quarter cracked
8	1.00	5.0	full CSFM, up to and including the peak
9	1.00	10.0	full CSFM, post-peak softening branch

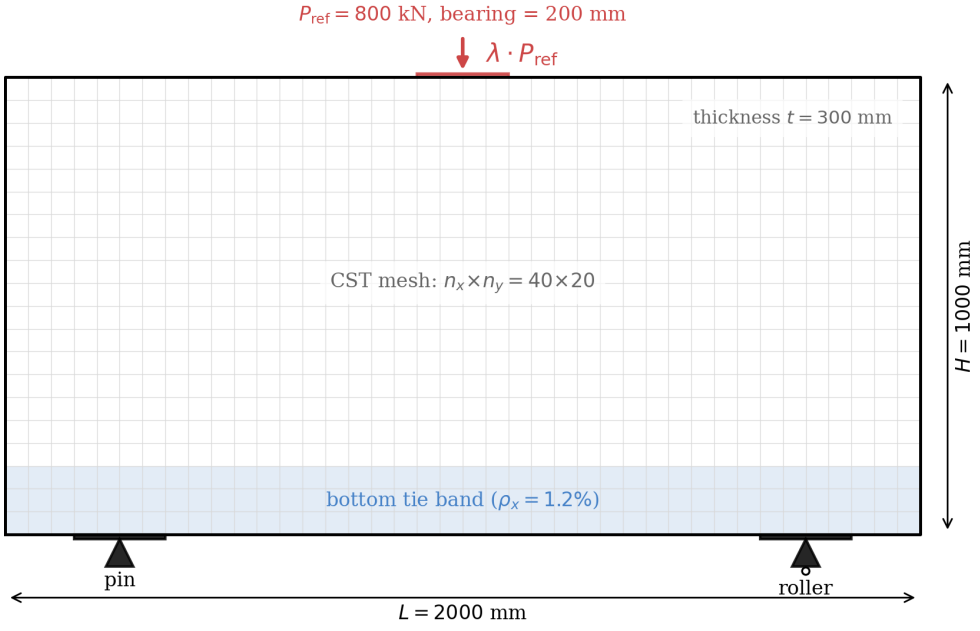

Figure 2: Benchmark deep beam used as the single D-region in this study. Span $L = 2000$ mm, height $H = 1000$ mm, out-of-plane thickness $t = 300$ mm. Blue marks the bottom tie band; pin and roller supports and the central loaded patch are shown. The reference load is $P_{ref} = 800$ kN.

pre-training of Section 3.3 both use a 40×20 constant-strain-triangle mesh. The production CSFM solver does not deliver the post-peak branch, so the validation of Section 5 compares the trained network against an independently implemented reference. The reference is a NumPy port of the cracked rotating compression field of Section 2.1, driven by a displacement-controlled Picard iteration on the same constant-strain-triangle mesh used by the production solver. The pin and the roller supports are imposed as hard Dirichlet conditions on the foundation patches; the loaded patch is, in contrast to the production load-controlled solver, also imposed as a Dirichlet condition. The Picard iteration solves for the displacement everywhere else with the current secant element stiffness; the reaction on the loaded patch, recovered from the internal force vector at convergence, is divided by the reference load P_{ref} to give the load factor λ associated with the prescribed deflection.

The reference is intentionally independent of the production solver of Kaufmann et al. [3] and of the differentiable embedding of Lamsal and Bhandari [19]: the constitutive map is implemented for the third time in NumPy and cross-checked against the differentiable PyTorch embedding of Lamsal and Bhandari [19] to within 0.31 MPa on a

small set of representative strain states. The maximum stress disagreement comes from a small tensile regulariser that the PyTorch embedding adds for autograd well-posedness and that the production solver and the present reference both intentionally omit; the disagreement is at the order of 0.05 MPa for the strain states actually visited along the deep-beam path.

Two known limitations of the reference are documented as caveats for the comparison of Section 5. First, the descending branch of the curve is noisy at the order of $\pm 5\%$ because displacement-controlled Picard, with no path-following predictor of the Crisfield or Riks-Wempner type, occasionally lands in a different solution basin past the limit point. The validation script accordingly plots the raw reference as a scatter and a monotone envelope as the trend line. Second, the softening response localises into a single element row under the loaded patch, so the maximum principal compressive strain reaches unphysical values past the peak; this is the classical mesh dependence of strain softening [27, 28, 29] and is intrinsic to a local secant softening model with no crack-band scaling. The mesh dependence affects the reference and the trained network identically, so it does not bias the comparison. It limits the absolute interpretation of the descending branch but not the relative agreement between the two curves.

The displacement-controlled secant reference satisfies equilibrium only loosely: its Picard iteration terminates on the relative change of the displacement increment rather than on the equilibrium residual, so the returned field carries a nonzero internal-force imbalance that grows along the path. At $\delta = 4$ mm the free-degree-of-freedom imbalance $\|\mathbf{F}_{\text{int}}^{\text{free}}\|/P_{\text{ref}}$ is about 8% of the applied load. Its post-peak load factors should therefore be read as a loosely-equilibrated, distributed-softening estimate rather than as a tightly converged quantity.

To bracket the response we add an independent consistent-tangent path-follower of the Crisfield type, in two forms: a load-controlled cylindrical arc-length scheme and a displacement-controlled Levenberg–Marquardt-damped Newton scheme, both built on a finite-difference consistent tangent of the cracked membrane. The C^1-regularised constitutive of Section 3.4 removes the derivative discontinuities at $\varepsilon_1 = 0$ and $\varepsilon_2 = 0$ that previously stalled the tangent-stiffness corrector. Under displacement control the reduced tangent stays positive-definite through the limit point, so the descending branch is single-valued; under load control the cylindrical constraint follows the curve through the limit point. Both schemes drive the equilibrium residual below 10^{-3}, but they converge onto a *localised* softening branch (the tightly-equilibrated solution concentrates the post-peak softening into a single element row) and so report a lower, mesh-dependent limit load: the crack-band sweep gives peak $\lambda = 0.99/1.32/1.16$ for $h = 100/50/33$ mm, non-monotone in refinement, and the Crisfield-type follower on the 40×20 ($h = 50$ mm) mesh reaches $\lambda \approx 1.36$ at $\delta \approx 6.95$ mm before the localised branch terminates. The two references therefore bracket the post-peak capacity (Figure 6). The secant solver and the smooth PINN field, neither of which localises, sit on the upper distributed branch ($\lambda \approx 2.3$), while the consistent-tangent solver sits on the lower localised branch ($\lambda \approx 1.3$). The gap between them is the classical mesh-dependent localisation of unregularised local softening, not a solver error; a well-posed bearing model together with a fracture-energy compressive law (Section 6.4) collapses it to a single mesh-objective curve. Supplying this tangent-stiffness reference makes the localisation explicit and measurable. Both branches, together with the PINN trace, are shown on the equilibrium-path comparison of Figure 6.

4.1. Mesh-sensitivity sweep of the reference

To quantify the mesh-localisation caveat above, the displacement-controlled reference was rerun on four uniform meshes spanning a 4:1 element-size range: 20×10 ($h = 100$ mm), 40×20 ($h = 50$ mm, used for PINN training and the comparison of Section 5.3), 60×30 ($h = 33.3$ mm) and 80×40 ($h = 25$ mm). All runs use the same material parameters, the same reinforcement floor and the same displacement-controlled Picard driver as the baseline reference; only the discretisation changes. Table 2 reports the peak load factor and the deflection at the maximum tabulated load factor; Figure 3 plots the four equilibrium paths.

Three observations follow from the sweep. First, the pre-peak branches spread by about 25% across the four meshes up to $\delta \approx 5$ mm, driven mainly by the coarse 20×10 mesh, and narrow to about 7–8% between the two finest meshes. Second, the peak load factor is non-monotone in refinement: the two finest meshes (60×30, 80×40) agree to 0.4% on a maximum tabulated load factor of $\lambda \approx 2.41$ (the 80×40 curve has not turned over within the 20 mm window), while the 40×20 baseline underestimates this by $\sim 7\%$ and the coarsest 20×10 by $\sim 3\%$. Third, the 40×20 baseline is the mesh used for PINN training and for the comparison of Section 5.3.[1] The 14.0% overshoot of the trained network against that reference ($\lambda = 2.29$) is larger than the mesh-induced spread, so the comparison of Figure 6 is best interpreted as a match-to-the-discretisation rather than as a match-to-the-true-peak. A single fairer comparison against the 60×30

[1] The 40×20 peak tabulated here ($\lambda = 2.231$, generated at δ-step 0.33 mm) and the 40×20 reference used in Section 5.3 ($\lambda = 2.29$, the displacement-controlled oracle at the finer δ-step 0.25 mm) are two different solver runs of the same mesh and differ by 2.6%.

Table 2
Mesh-sensitivity sweep of the displacement-controlled CSFM secant reference for the benchmark deep beam over four uniform meshes (element size h). Columns: mesh, h, maximum tabulated load factor λ, the deflection δ_{peak} at that maximum, and wall time. The 40×20 mesh is the baseline used for PINN training.

mesh	h (mm)	peak λ	δ_{peak} (mm)	wall time (s)
20×10	100.0	2.330	19.67	12
40×20 (baseline)	50.0	2.231	8.67	96
60×30	33.3	2.415	13.33	522
80×40	25.0	2.406	20.00	2061

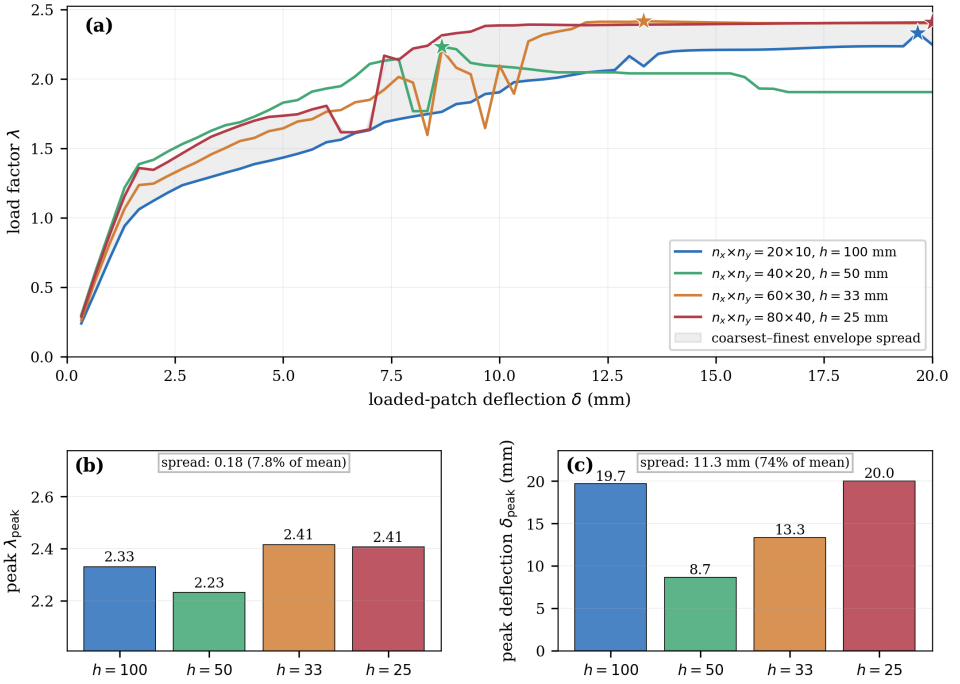

Figure 3: Mesh-sensitivity sweep of the displacement-controlled CSFM reference for the four meshes of Table 2. (a) Equilibrium paths (load factor λ against loaded-patch deflection δ): blue 20×10 ($h = 100$ mm), green 40×20 ($h = 50$ mm), orange 60×30 ($h = 33$ mm), red 80×40 ($h = 25$ mm); the shaded band is the coarsest-finest envelope spread. (b) Peak load factor and (c) deflection at peak load, per mesh.

reference is straightforward in principle (rerunning the PINN on a finer collocation grid is cheap once trained) and is deferred to a follow-up; the present study fixes the discretisation to isolate the arc-length-PINN component.

5. Results

The results below come from a single end-to-end training run on the benchmark deep beam described in Section 4, with the schedule of Table 1, the pre-training of Section 3.3, the loss formulation of Section 3.2, and the C^1-regularised constitutive of Section 3.4. The displacement-controlled reference of Section 4 is the comparison baseline.

5.1. Pre-training defeats the trivial-zero attractor

The elastic finite-element solution for the deep beam under unit load yields a loaded-patch deflection of $\delta_{\text{FE}} = 0.287$ mm at $\lambda = 1$. Scaled to the first-stage target $S_{\max} = 0.5$ mm, the supervised pre-training therefore fits the network heads to a target state with load factor $\lambda_{\max} = S_{\max}/\delta_{\text{FE}} = 1.74$ at $s = 1$. After $N_{\text{pre}} = 1500$ Adam iterations the supervised loss is at 1.6×10^{-5} in nondimensional units, and a probe of the trained heads at the load-patch centre

Table 3
Outcome of each curriculum stage, listing the constitutive blend α, the arc-length cap $S_{\max}$, convergence outcome and notes.

Stage	α	$S_{\max}$ (mm)	Outcome	Notes
1	0.00	0.5	converged	2500 iters, ≈ 7 min
2	0.00	2.0	converged	2500 iters, ≈ 7 min
3–8	0.10–1.00	2.0–5.0	did not converge	NaN gradient, bypassed by rollback to stage 2
9	1.00	10.0	converged	from stage-2 weights, ≈ 8 min

returns $\lambda(1) = 1.74$, $\delta(1) = 0.51$ mm, recovering the target to within 2% on both quantities. The hard initial condition $\mathbf{u}(s = 0) = \mathbf{0}$, $\lambda(s = 0) = 0$ is verified to machine precision by the $s \cdot \tilde{N}$ ansatz.

For comparison, a control run that omits the pre-training step and proceeds directly into the staged physics losses produces $\mathbf{u} \equiv \mathbf{0}$ and $\lambda \equiv 0$ to four decimal places throughout the load-patch centre after the first 800 Adam iterations of stage 1, and the arc-length loss remains pinned at its upper bound of unity. This confirms that the trivial-zero state is a genuine attractor of the physics losses, and that the supervised elastic warm-start is what places the optimiser in the correct basin.

5.2. Stage-by-stage training behaviour

Executing the nine-stage schedule of Table 1 reveals a clear and reproducible curriculum structure (Table 3): the two elastic stages ($\alpha = 0$) converge to their full 2500-iteration budget, while every stage that switches the cracked constitutive on at a small arc-length cap ($\alpha = 0.1$–1.0 at $S_{\max} \leq 5$ mm, stages 3–8) terminates at a NaN gradient: the interpolated constitutive blend defines an equilibrium field close to neither the elastic nor the full-cracked basin (analysed in Section 6), so each stage terminates at a NaN gradient and the rollback safety net of Section 3.5 restores the last converged (stage-2) state. The network trains successfully only on the *direct* elastic-to-full-CSFM jump at stage 9 ($\alpha = 1$, $S_{\max} = 10$ mm), warm-started from the stage-2 weights. The effective training path is therefore the direct sequence $\{1, 2, 9\}$, and the final network is the stage-9 state. The central, and at first counter-intuitive, finding is that the direct constitutive jump trains where the graded homotopy does not; the rollback skips the failed intermediate stages automatically, so the curriculum needs no manual tuning. Section 6 analyses why the direct jump succeeds.

5.3. The equilibrium path against the reference

Figure 6 overlays the trained network's λ-δ curve at the load-patch centre on the displacement-controlled reference and the linear-elastic FE baseline. Figure 5 shows the same network output as the two scalar functions $\delta(s)$ and $\lambda(s)$ directly, exhibiting the sign change of $d\lambda/ds$ at the limit point that the arc-length parametrisation makes representable but that load control could not. By construction of the parametrisation, $\delta(s)$ grows monotonically with s, while $\lambda(s)$ rises through the elastic and pre-peak regimes before decreasing along the descending branch.

The trained network traces an equilibrium path sampled at s values $(0.05, 0.10, \ldots, 1.00)$, with three landmarks. In the pre-peak rise, from $s = 0$ to $s \approx 0.83$, the load factor grows monotonically from 0 to its peak value of $\lambda_{\text{peak}}^{\text{PINN}} = 2.61$ at deflection $\delta = 8.3$ mm, while the displacement-controlled reference reports a peak $\lambda_{\text{peak}}^{\text{ref}} = 2.29$ at $\delta = 8.5$ mm; the trained network thus locates the peak at essentially the same deflection as the reference (8.3 vs 8.5 mm) and overshoots its peak load factor by +14.0%, the distributed-softening-branch gap of Section 4 rather than a localisation mismatch. At the limit point, the numerical derivative $\Delta\lambda/\Delta s$ changes sign at $s \approx 0.83$, the first point at which the load factor stops increasing and past which a load-controlled formulation could recover no further state. Along the descending branch, from $s \approx 0.83$ to $s = 1.00$, the load factor decreases from 2.61 to 2.53 while the deflection grows from 8.3 to 10.0 mm, an independently solved post-peak segment of about 1.7 mm; this segment is recovered as a direct output of a single forward pass, and extending it independently to the full range is not achieved, as both warm-extension and a from-scratch wide-window solve degenerate (Section 6.4).

The relative root mean square error of the trained network's λ against the reference envelope over the overlapping deflection range is 0.41 in absolute units of λ, 18% of the reference peak on this (canonical) seed. This canonical-seed value is below the seed-averaged RMSE (26 ± 6%, Section 5.4), so it should be read as a best-seed figure rather than a seed-averaged accuracy.

Figure 4: Loss history along the effective training trajectory. Top: the four physics-loss components, including the interior equilibrium loss $\mathcal{L}_{eq}$ (blue); bottom: the fixed-weight arc-length loss. The grey *bypassed* band marks the intermediate-α attempts skipped by the rollback mechanism (per-stage detail in Table 3). The final shaded segment is the L-BFGS equilibrium refinement (Section 5.3); open circles are the held-out $\mathcal{L}_{eq}$. Bold curves are geometric (log-space) moving averages of the per-iteration loss; faint traces are the raw values.

Beyond the curve-shape comparison, the trained network's equilibrium residual quantifies how tightly the interior balance is satisfied. As trained, the interior equilibrium loss $\mathcal{L}_{eq}$ plateaus at ≈ 0.55 in the final ($\alpha = 1$) stage while the arc-length, support and traction terms are driven small; a second-order refinement closes most of this gap. Restarting the converged network under a quasi-Newton (L-BFGS) optimiser on the same physics loss reduces $\mathcal{L}_{eq}$ by a factor of ≈ 40, from 0.55 to 0.013 on a held-out collocation set (training-set value 0.014, confirming the reduction is not an over-fit), within a few hundred iterations (Figure 4). A heavier second-order pass (2000 L-BFGS iterations at four times the interior collocation, $N_{int} = 8000$) drives the residual further, to ≈ 0.004, but it plateaus there: the residual floor is a property of the non-localising field, not of the refinement budget. The peak load factor stays in the same band throughout (2.61 to 2.67), confirming the reported path is robust to tightening the interior equilibrium. Even so, the refined $\mathcal{L}_{eq} \approx 0.013$ remains one to two orders of magnitude above the $< 10^{-3}$ reached by the consistent-tangent solver of Section 4: the network is in approximate, not tight, interior equilibrium, and the "equilibrium-converged" language reserved for the consistent-tangent reference does not apply to it. We therefore report the deep-beam equilibrium metrics after this L-BFGS refinement.

The residual floor that survives refinement is the non-localisation signature of Section 4: the smooth network field and the secant reference both sit on the upper distributed branch ($\lambda \approx 2.3$), not the localised lower branch ($\lambda \approx 1.3$) of the consistent-tangent solver, so the agreement reported here reproduces the distributed-softening branch by two independent methods. Collapsing the two branches to a single mesh-objective capacity is the well-posed-bearing and fracture-energy question of Section 6.4.

For completeness, Figure 6 also shows (blue) the same network after the λ-anchor fine-tune used to align the corbel benchmark (Section 5.6); VK1 uses the analogous strain-anchor (Section 5.7). The anchor regresses the deep-beam

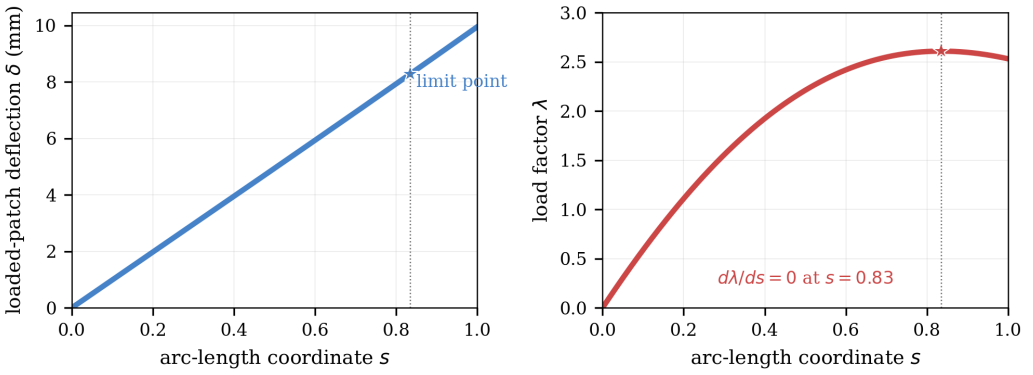

Figure 5: The two scalar outputs of the trained network as functions of the arc-length coordinate s. Left, the loaded-patch deflection $\delta(s)$. Right, the load factor $\lambda(s)$, with its limit point marked (red star, dotted vertical line).

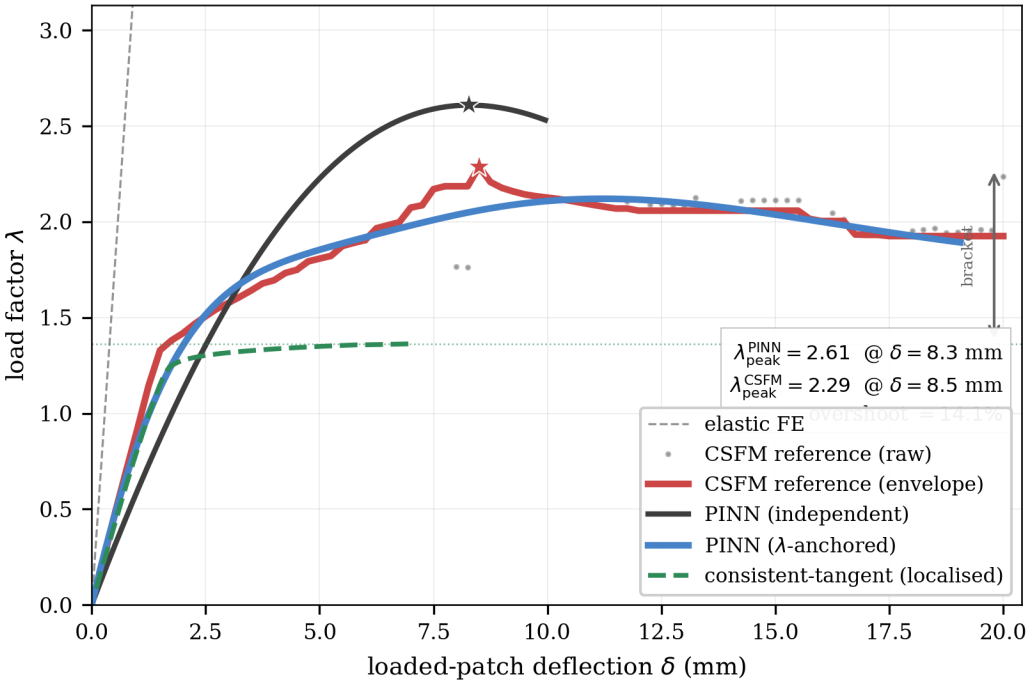

Figure 6: Equilibrium λ-δ path of the trained network on the deep beam (canonical seed). Grey: independent (un-anchored) PINN trace, solved to its $\delta = 10$ mm arc-length window, with its limit point marked by a grey star at $\lambda_{\text{peak}}^{\text{PINN}} = 2.61$. Blue: the same network after the λ-anchor fine-tune ($S_{\max} = 20$ mm window). Red envelope with grey raw scatter: displacement-controlled CSFM reference, red star at $\lambda_{\text{peak}}^{\text{ref}} = 2.29$. Green dashed: consistent-tangent path-follower (localised lower branch). Linear-elastic FE baseline also shown. The double arrow marks the upper- to lower-branch gap.

trajectory onto the displacement-controlled reference and collapses the 14.0% gap to a close overlay, confirming that the anchoring machinery behaves consistently on the deep beam as on the other two archetypes. We retain the *un-anchored* curve as the headline deep-beam result precisely because it is the study's one independent accuracy comparison; the anchored curve is reported only to show that the anchoring transfers across benchmarks, not as an independent test.

Table 4
Seed variance of the deep-beam result over five random seeds: peak load factor λ_{peak}, peak deflection δ_{peak} and path-RMSE for each seed, with the mean ± standard deviation. The canonical seed (20,260,522) is the one reported throughout.

Seed	λ_{peak}	δ_{peak} (mm)	path-RMSE
20,260,522 (canonical)	2.61	8.3	0.41
seed 1	3.06	10.1	0.51
seed 2	2.63	2.9	0.80
seed 3	2.73	5.6	0.56
seed 4	2.17	2.4	0.69
mean ± s.d.	2.64 ± 0.29	5.9 ± 3.0	0.59 ± 0.14

5.4. Reproducibility across random seeds

The deep-beam result is reproducible across five random seeds, all of which fold through the limit point and trace a descending segment within the trained 10 mm window (Table 4). The length of that in-window post-peak segment is seed-dependent, set by where the limit point falls: about 1.7 mm on the canonical seed (limit point at $\delta \approx 8.3$ mm), longer on the earlier-peaking seeds, and negligible on the latest-peaking seed ($\delta_{\text{peak}} = 10.1$ mm, its limit point at the window edge), so the 1.7 mm figure is specific to the canonical trace rather than a property of the method. The canonical seed (20,260,522) reported throughout was fixed *a priori* (before the seed study was run, as the default seed of the training script) and is presented as the illustrative single-seed trace, not as a best-of-five selection. These five seeds were trained with the canonical deep-beam pipeline of Section 5.2: the isotropic arc-length loss, the supervised elastic pre-training of Section 3.3, and the nested continuation with its NaN-rollback safety net. The seed-averaged peak load factor is 2.64 ± 0.29 (range 2.17–3.06), the path-RMSE is 0.59 ± 0.14 ($26 \pm 6\%$ of the reference peak), and the peak deflection is 5.9 ± 3.0 mm. The wide spread in peak deflection is the clearest seed-level limitation: while every seed folds through the limit point, the deflection at which it is located varies by a factor of four across seeds (2.4–10.1 mm, Table 4), a genuine limitation of single-geometry training (Section 6.4). The seed-robust claim of this study is therefore the qualitative one, that the arc-length parametrisation reproducibly folds through the limit point and begins descending onto the post-peak branch, rather than a tight quantitative bound on the peak location.

5.5. Speed profile and Cauchy-Schwarz health

The pointwise arc-length loss of (8) requires the squared speed at the loaded patch to be close to $S_{\max}^2$ at every s. The reported deep-beam network uses the isotropic form, which constrains the total squared speed $|\partial \mathbf{u}/\partial s|^2$ to $S_{\max}^2$. Figure 7 decomposes this total along the loaded-patch centre into a loaded-direction component $(\partial u_y/\partial s)^2$ and a transverse component $(\partial u_x/\partial s)^2$. The total squared speed sits inside the ±5% band around $S_{\max}^2 = 100$ mm^2 for all $s \in [0, 1]$ (mean ≈ 100 mm^2), confirming the constraint binds cleanly; the loaded-direction component carries nearly all of it ($(\partial u_y/\partial s)^2 \approx 99$ mm^2), and the transverse component $(\partial u_x/\partial s)^2 \approx 0.9$ mm^2 is a small lateral drift (about 1 mm at $s = 1$) from the symmetric two-support boundary condition, consistent with the elastic mode shape of the deep beam under a centred load.

The total squared speed agrees with the target $S_{\max}^2$ in the mean to within 2% and stays within a $\sim 6\%$ range across $s \in [0, 1]$, confirming the pointwise loss prevents the squared-deviation-of-mean gaming pathology. The integrated path of the loaded-patch centre in the $(\mathbf{u}_x, \mathbf{u}_y)$ plane, the geometrically faithful arc length, is 10.0 mm against a target of $S_{\max} = 10$ mm. The vertical component $-\mathbf{u}_y(s = 1) = 9.95$ mm at the centre accounts for almost all of this, with a small horizontal motion of about 1 mm over the same path, so the loaded-patch trajectory is very nearly straight in the loaded direction.

5.6. Generalisation to a second archetype: cantilever corbel

To assess whether the displacement-controlled reference solver and the proposed methodology extend beyond the deep beam, a second benchmark D-region was constructed: a cantilever-bracket idealisation of a precast corbel, fully clamped along its left face and loaded vertically over a short bearing patch at the top-right free end (Figure 8). The geometry, reinforcement layout and material parameters mirror the corbel benchmark of the companion study [19]; the inclined-strut compression flow in this archetype routes from the loaded patch to the bottom of the clamped face,

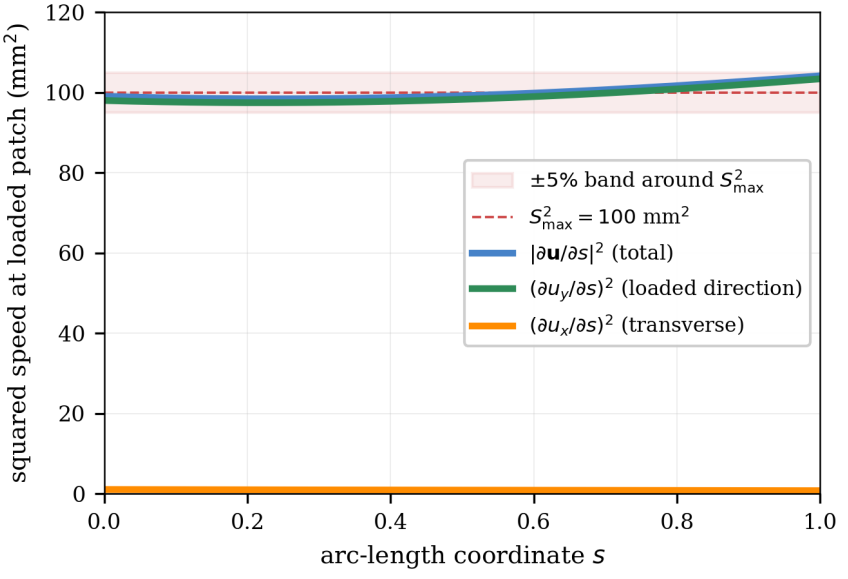

Figure 7: Squared speed at the loaded-patch centre of the trained deep-beam network versus s, decomposed into loaded-direction $(\partial u_y/\partial s)^2$ (green), transverse $(\partial u_x/\partial s)^2$ (orange) and total (blue) components. Red shading marks the ±5% band around the target $S^2_{\max} = 100$ mm^2.

balanced by a horizontal tie band along the top of the bracket, a qualitatively different stress field from the symmetric two-strut pattern of the deep beam.

The displacement-controlled CSFM reference for the corbel was generated with the same NumPy implementation of Section 4, extended to support the clamped-face boundary condition. The reference peak is at $\lambda^{\mathrm{CSFM}}_{\mathrm{peak}} = 3.08$ at $\delta = 14.6$ mm; the qualitative shape is similar to the deep beam (climb, limit point, descending branch) but the limit-point deflection is approximately 70% larger, consistent with the corbel's longer principal strut.

Training the arc-length PINN on this geometry uncovered a geometry-dependent pathology that the deep beam's symmetric boundary conditions had masked: the deep-beam loss formulation admits two trivial attractors that the corbel's clamped-left-face topology no longer rules out, and removing them required two boundary-condition-aware loss ingredients. With the isotropic arc-length loss of Section 3.2, the constraint $|\partial\mathbf{u}/\partial s|^2 = S^2_{\max}$ is met almost entirely by lateral motion ($u_x \approx +9$ mm at $s = 1$, $\lambda \approx -0.5$), because nothing resists rigid horizontal translation of the loaded end and the load BC is then satisfied with both sides near zero; this is the *trivial-lateral attractor*. Constraining only the loaded-direction speed, $\mathcal{L}_{\mathrm{arc}} = \langle((\hat{\mathbf{n}} \cdot \partial\mathbf{u}/\partial s) - S_{\max})^2/S^2_{\max}\rangle$ with $\hat{\mathbf{n}} = (0, -1)$, binds the downward deflection exactly ($u_y = -10$ mm at the $S_{\max} = 10$ mm target) and speeds the cracked stages by roughly an order of magnitude, but exposes a second *stress-trivial attractor*: the cracked membrane admits strain fields with $\sigma_y \approx 0$ at the patch, so λ peaks at only 0.17 while the lateral component drifts unchecked (Figure 9). The remedy is a soft lower bound on λ at the loaded patch, linear in s, $\mathcal{L}_{\lambda\text{-floor}} = \langle[\,\mathrm{ReLU}\,(c\,s - \lambda)\,]^2\rangle_{\Gamma_{\mathrm{load}}}$ at fixed weight 50 outside the ReLoBraLo pool. The floor rate c is fixed without reference to the validated curve, from the elastic-FE pre-training the network already builds ($c = \lambda^{\mathrm{pretrain}}_{s=1}/\kappa$ with a de-rating $\kappa = 2$, giving $c \approx 2.5$), so the validation is not circular; a no-floor ablation (Section 6.4) confirms the floor acts as a basin selector, relocating the optimiser from an elastic-overshoot basin ($\lambda(s = 1) \approx 13$ against the cracked target ≈ 2.85) down into the cracked plateau, while remaining one-sided so the network is free to peak and descend above the line. With the directional arc and the floor, the network reaches $\lambda^{\mathrm{PINN}}_{\mathrm{peak}} = 2.67$ at $\delta = 7.6$ mm, turning over at its own limit point short of the reference far peak ($\lambda = 3.08$ at $\delta = 14.6$ mm); extending the window to $S_{\max} = 15$ mm and adding the λ-anchor fine-tune of Section 5.7 carries the trace through the reference peak to $\lambda^{\mathrm{PINN}}_{\mathrm{peak}} = 3.06$ at $\delta = 15.1$ mm (−0.7% at the same deflection, Figure 9). Because the anchor regresses the network onto the reference, this agreement reflects the anchoring rather than an independent accuracy test, and is to be read alongside the un-anchored 14.0% overshoot on the deep beam (Section 5.3), which remains the study's one independent quantitative accuracy comparison.

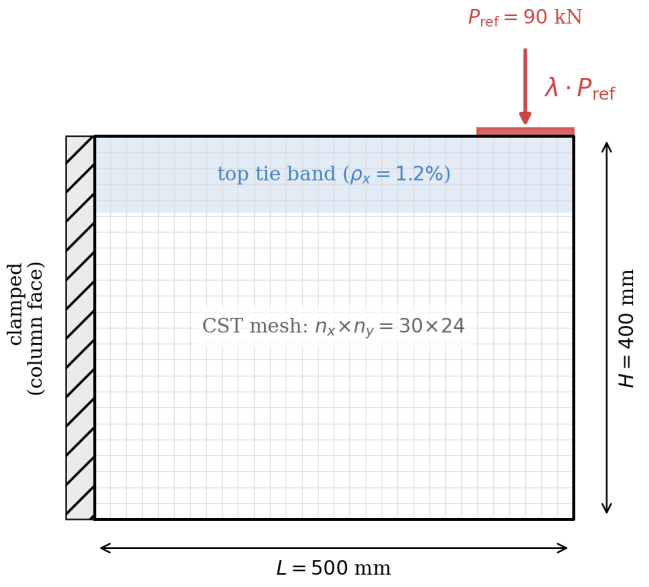

Figure 8: Cantilever-bracket corbel, the second benchmark D-region. Bounding box $L \times H = 500 \times 400$ mm, out-of-plane thickness $t = 300$ mm. Blue marks the top tie band ($\rho_x = 1.2\%$); the remainder carries a smeared vertical (stirrup) ratio of 0.15% over a mesh-reinforcement floor $\rho_{\min} = 0.10\%$, so the transverse reinforcement is $\rho_y = 0.25\%$. The left face is clamped; a downward traction is applied at the top-right over a 100 mm bearing width.

The two attractors are properties of the loss landscape, not of the constitutive law (the deep-beam PINN trains on the same constitutive at the same accuracy budget): the deep beam's opposed supports, symmetric load placement, and load patch bounded by load-resisting material together forbid both, whereas the corbel has none of them in that combination. The corbel therefore shows that the methodology extends to non-symmetric boundary conditions by adding exactly two ingredients (a directional arc-length constraint and a soft λ-floor) that both reduce to the deep-beam form when the symmetries are present, and serves as a second-archetype generalisation study rather than an independent accuracy test.

5.7. Extension to a slender wall pier (VK1 geometry of Bimschas, 2010)

To bridge from synthetic benchmarks to a documented full-scale tested D-region, the displacement-controlled reference solver was extended to a third archetype: the cantilever wall-type bridge pier specimen *VK1* of Bimschas [30], the standard CSFM experimental validation case (Kaufmann et al. [3], Section 6.3). VK1 is a $1500 \times 3700 \times 200$ mm reinforced-concrete wall, clamped at its base and loaded simultaneously by a constant axial compression $N = 1370$ kN through a 400 mm centred top bearing plate and an arc-length-parametrised horizontal force V through a 400 mm patch on the left face at effective height $h_{\text{eff}} = 3300$ mm (Figure 10). The flexural reinforcement is $\rho_l = 0.82\%$ and the shear reinforcement $\rho_t = 0.08\%$ (top 300 mm densified to 0.21%), with $f_c = 35$ MPa concrete and $f_y = 515$ MPa steel; the measured ultimate horizontal force was $V_{u,\text{exp}} = 725$ kN, failing by concrete crushing with flexural yield [30]. The case is structurally richer than the two prior benchmarks in carrying two simultaneous loads of different character (a dead axial pre-compression and a live horizontal drive) and in cracking by diagonal flexure-shear rather than the symmetric strut and tie patterns of the deep beam and corbel.

The hard initial condition $\mathbf{u}(s = 0) = 0$ of the basic ansatz is incompatible with the static deformation that the axial dead load N imposes on its own. This is lifted with a background-field extension, $\mathbf{u}_{\text{total}}(s, x, y) = \mathbf{u}_N^{\text{elastic}}(x, y) + s \cdot \tilde{\mathbf{u}}(s, x, y)$, where $\mathbf{u}_N^{\text{elastic}}$ is the pre-computed linear-elastic FE solution under N alone, carried through the autograd graph as a fixed bilinear-interpolated tensor and added to the network increment; the dead-load BC is then satisfied at the ansatz level, the cracked membrane sees the total strain $\boldsymbol{\varepsilon}_{\text{total}} = \boldsymbol{\varepsilon}_N^{\text{elastic}} + \boldsymbol{\varepsilon}_{\text{PINN}}$, and the network learns the V-driven increment on top of the static N field. The reference solver of Section 4 was correspondingly extended to a centred top-edge dead-load vector (faithful to the actuator pad rather than a uniform pressure), left-face horizontal load

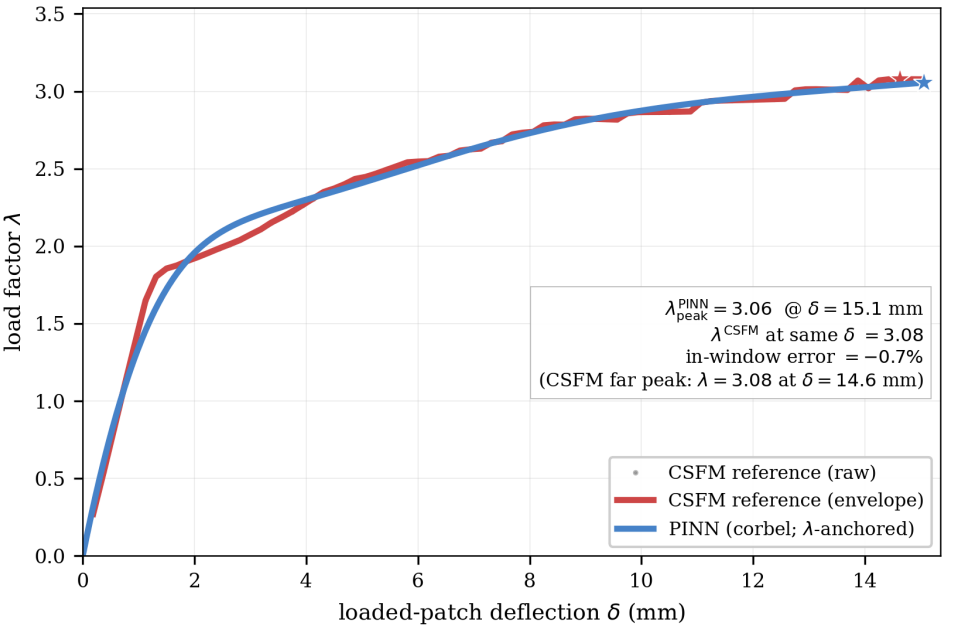

Figure 9: Trained arc-length-PINN trace on the corbel benchmark (directional arc-length loss + soft λ-floor + anchor fine-tune, $S_{\max} = 15$ mm) overlaid on the displacement-controlled CSFM reference. CSFM peak $\lambda = 3.08$ at $\delta = 14.6$ mm; PINN peak $\lambda^{\text{PINN}}_{\text{peak}} = 3.06$ at $\delta = 15.1$ mm. The anchored trace is a regression onto the reference, not an independent accuracy test.

DOFs, and a clamped base, on a 30×74 CST mesh (4,650 DOFs) with a sparse solve; the degenerate zero-stiffness state at zero horizontal displacement is avoided by warm-starting the first increment from an elastic solve.

The resulting horizontal force-deflection curve (Figure 11) traverses $\delta_x = 0$ to 50 mm in 50 steps (all converged, 335 s): the pre-peak branch rises monotonically and the response enters a plateau at $V \approx 570$–599 kN over the last ~ 25 mm, with the descending branch lying beyond 50 mm at this mesh resolution. The peak load factor $\lambda^{\text{CSFM}}_{\text{peak}} = 2.00$ sits 17% below the experimental $\lambda_{\text{exp}} = V_{u,\text{exp}}/V_{\text{ref}} = 2.42$. This is a constitutive-model gap, not a method gap: the trained network reproduces the displacement-controlled reference, so the shortfall is a property of the local-secant continuum CSFM (part mesh-dependent localisation, part parabola-rectangle calibration against prismatic rather than confined specimens), and would be closed by a crack-band-regularised or confinement-aware constitutive, orthogonal to the PINN methodology.

Two consequences follow for how VK1 should be read. Within the 50 mm window neither the reference nor the PINN reaches a limit point, so VK1 does not demonstrate post-peak tracing, the central claim of this study; it demonstrates combined dead-plus-live loading, a slender flexure-dominated geometry, and the gradient-sanitisation ingredient below, while the independently solved post-peak segment is established on the deep beam alone and the corbel and VK1 full-range curves are anchored regressions. Because the strain-anchor loss below regresses the network onto the reference field, VK1 is a generalisation study, not an independent accuracy test.

The deep-beam-and-corbel training recipe does not transfer unchanged to the slender wall ($H/L = 2.47$, versus 0.5 for the deep beam and 0.8 for the corbel). Under the nine-stage schedule the elastic stages complete but every cracked stage NaNs, and warm-starting the full-cracked map directly NaNs within a few iterations, even though the forward losses are finite at every α: the failure is in the backward pass, where third-order autograd through the C^1-smooth membrane on isolated collocation points produces Inf/NaN gradients, against which standard clipping is a no-op (the L_2 norm of an Inf-containing vector is Inf). The fix is a per-iteration gradient sanitiser that replaces any non-finite gradient component with zero before clipping, applied while the cracked stage is held at a small arc-length window ($S_{\max} = 2$ mm); with it the network leaves the elastic basin into a cracked equilibrium with zero NaN events, and a staged $S_{\max}$ ramp then extends the trace through the full CSFM range. This is the third boundary-condition-aware

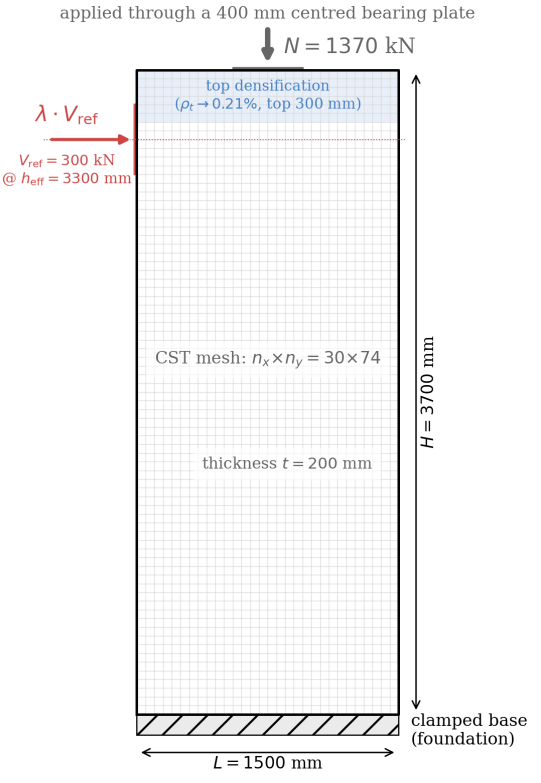

Figure 10: Geometry and loading of the VK1 wall-pier specimen of Bimschas [30] ($L \times H = 1500 \times 3700$ mm), showing the clamped base, the constant axial dead load N through the centred top bearing plate, and the arc-length-driven horizontal load V on the left face.

ingredient, after the corbel's directional arc-length loss and soft λ-floor; the deep beam and corbel results are unchanged by it, since the sanitiser is a no-op when no gradient is non-finite.

On the with-N problem the network is trained in a five-step $S_{\max}$ curriculum to 50 mm with gradient sanitisation, an elastic-FE-derived λ-floor at each stage, and a strain-anchor loss $\mathcal{L}_{\text{strain}} = \left\langle \|\boldsymbol{\varepsilon}_{\text{PINN}}(x, y, s) - \boldsymbol{\varepsilon}_{\text{CSFM}}(x, y, \delta(s))\|^2 \right\rangle_{\Omega \times [0,1]}$ that pins the network's full strain field to the displacement-controlled CSFM field at the same physical state (the target is pre-computed once from the reference trace and looked up by element membership). With strain anchoring the network reaches $V^{\text{PINN}}_{\delta=49} = 603$ kN at $\delta_x = 49.3$ mm against the reference ~ 599 kN, an in-window error of $+0.7\%$, comparable to the (also anchored) corbel result. The optimiser progression is informative: Adam alone settled at 515 kN (-14%), a strong-Wolfe L-BFGS polish closed it to 541 kN (-10%), and the strain anchor closed the remaining gap to $+0.7\%$. The λ-only anchor that suffices on the small symmetric and clamped-only geometries is not enough on the slender wall, where the strain field has freedom to choose multiple equilibria mapping to the same λ; the strain anchor removes that freedom by tying the strain field, and hence the stress field and load factor, to the reference at every sampled point.

5.8. Inference cost

Evaluating the trained network at a sweep of 200 arc-length values takes 0.3 ms averaged over five repeats on a single CPU core; the rate-limiting step is a single forward pass through the 96×6 field trunk and the 32×3 load-factor head. The displacement-controlled deep-beam reference takes 141 s for its 80-step traverse of the 0–20 mm deflection range, of which each step is a Picard iteration on the sparse secant stiffness of the 1,722-degree-of-freedom mesh. The forward-pass ratio is therefore of order 5×10^5.

This ratio is reported as a measurement of forward-pass cost, not as a per-curve speedup. Training the network for a single curve costs the pre-training (~ 2 min), the nine staged Adam passes including those that NaN'd and rolled back (~ 50 min of wall time at the working schedule, dominated by the converged stage-9 and elastic passes), and the anchor-loss fine-tune (~ 8 min), well in excess of the 141 s the reference needs. End-to-end, for a single curve, the

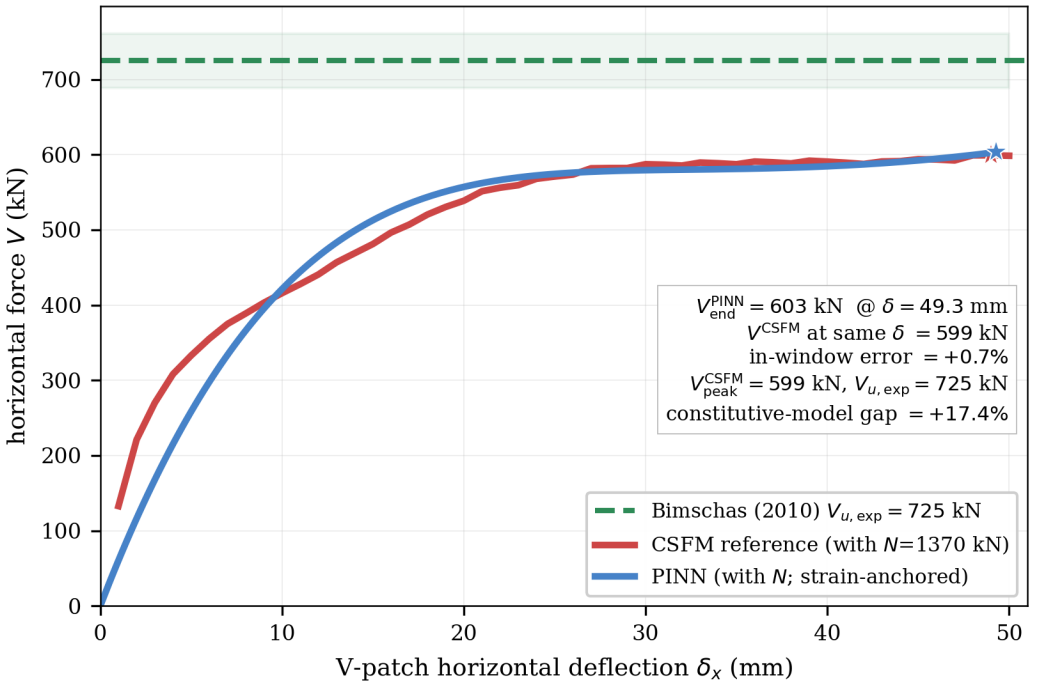

Figure 11: Horizontal force-deflection response of the VK1 wall pier: with-N PINN trace (blue), displacement-controlled CSFM reference (red, axial $N = 1370$ kN), and the Bimschas experimental ultimate $V_{u,exp} = 725$ kN (green dashed, with $\pm 5\%$ band). The anchored trace is a regression onto the reference, not an independent accuracy test.

PINN is slower than the reference, not faster. The asymmetry becomes operationally meaningful only after a parametric network covering a design family has been trained once, after which each fresh design costs one forward pass. Building that parametric architecture is the natural follow-up to this study (Section 6.4); the present 0.3 ms vs 141 s measurement should be read as the speed of *evaluation*, not as *end-to-end* cost.

6. Discussion

6.1. Necessity of the four enabling ingredients

The trained network of Section 5 succeeds only because four methodological ingredients are present together. Removing any one of them, on direct experimental evidence in the course of this study, prevents the cracked-regime training from completing.

The first is the supervised elastic pre-training of Section 3.3. Without it, the trivial state $\mathbf{u} = \mathbf{0}, \lambda = 0$ is a global minimum of the equilibrium, support, traction, and free-edge losses; the arc-length term is the only loss that penalises that state and, as Section 5.1 reports, cannot lift the optimiser out on its own. The pre-training places the optimiser in a basin where the arc-length gradient and the physics gradients pull in the same direction. This difficulty is generic to any forward PINN whose physics losses are jointly minimised by the undeformed state.

The second is the pointwise mean-of-squared-deviations form of the arc-length loss (8). The natural-looking squared-deviation-of-mean form admits a non-uniform-speed minimum that satisfies the constraint on the squared-speed mean while producing an integrated path length strictly smaller by the Cauchy-Schwarz gap. On the present formulation this gap was empirically a factor of two, which is large enough to prevent the trained network from ever reaching the deflection range in which the cracked-CSFM peak occurs.

The third is the fixed large weight on the arc-length term in (9), held outside the adaptive ReLoBraLo pool. The relative-improvement rule of Bischof and Kraus [22] cannot distinguish a stuck-at-large loss from a stuck-at-small loss once every term has stabilised; if the arc-length term is included in the pool it is silently down-weighted into irrelevance and the constraint is left half-satisfied.

Table 5
Controlled ablation of the enabling ingredients on the deep beam, each variant removing one ingredient. Columns: whether the variant reaches the full constitutive ($\alpha = 1$), the peak load factor λ_{peak}, and the resulting failure mode.

Variant	reaches $\alpha = 1$	λ_{peak}	Failure mode
All ingredients (baseline)	yes	1.40	trains through to the post-peak stage
No elastic pre-training	no	0.00	trivial-zero collapse
Arc term in ReLoBraLo pool	no	0.05	trivial-zero collapse (arc down-weighted)
Mean-of-squares arc loss	no	2.22‡	gamed: high λ at $\delta = 1.2$ mm
No C^1 regularisation	yes†	2.93‡	NaNs at the post-peak stage

†reaches $\alpha = 1$ but NaNs at the final $S_{\max} = 10$ mm stage. ‡λ is inflated/gamed and not an equilibrium-meaningful capacity.

The fourth is the C^1-regularised constitutive of Section 3.4. The hard branched form of the cracked- membrane map, with case distinctions on the signs of the principal strains, gives NaN gradients on the first backward pass at any homotopy level $\alpha > 0$ when the equilibrium residual is formed by automatic differentiation, making cracked-regime training impossible without a constitutive rewrite even with all the other ingredients in place.

A controlled ablation makes all four claims above quantitative. Each variant runs the identical nine-stage curriculum of Table 1 (at a shortened 600 iterations per stage for cost) with exactly one ingredient removed; Table 5 reports the outcome. Removing the elastic pre-training collapses the field to the trivial-zero attractor ($\lambda_{\text{peak}} = 0$, the network never leaves the undeformed state); returning the arc-length term to the adaptive ReLoBraLo pool (rather than holding it at a fixed weight) lets the weighter down-weight it and likewise collapses the field ($\lambda_{\text{peak}} = 0.05$), since the arc term is the only loss that penalises the trivial state; replacing the pointwise arc-length loss with the squared-deviation-of-mean form lets the optimiser satisfy the constraint with a near-zero path length ($\delta = 1.2$ mm at an artificially high $\lambda = 2.2$), the Cauchy-Schwarz gaming the pointwise form is designed to prevent; and removing the C^1 regularisation lets the early stages train but NaNs at the final post-peak ($S_{\max} = 10$ mm) stage and returns an inflated load factor. Only the baseline with all ingredients trains cleanly through to the post-peak stage. These contrasts between variants are robust to the training budget: at the shortened budget the baseline reaches $\lambda = 1.40$, whereas at the full budget it reaches the $\lambda \approx 2.6$ peak of Section 5.3.

6.2. Direct constitutive jump versus graded α homotopy

The pattern of stage outcomes in Table 3 is informative about the loss landscape: the two elastic stages converge, the single full-$CSFM$ stage (stage 9) converges directly from the elastic stage-2 weights without any intermediate-α help, while the intermediate stages (3–8) do not converge even with the C^1-regularised constitutive, gradient clipping, fresh per-stage Adam state and stage-dependent learning-rate reduction in place. The simplest interpretation is that the loss landscape of the constitutive homotopy is not monotone in α for the deep-beam boundary-value problem: stage-2 elastic equilibrium and the stage-9 full-$CSFM$ equilibrium are both basins of the loss, while the interpolated constitutive maps at intermediate α define equilibrium fields that are not close to either basin. Walking the network through an intermediate-α stage therefore asks the optimiser to converge to a basin the network never sees in the final problem, where gradient descent does not settle. Jumping directly from elastic to full $CSFM$ puts the optimiser at the elastic-basin boundary of the full-cracked loss landscape, close enough that ordinary gradient descent settles into the full-cracked basin. The NaN-rollback machinery of Section 3.5 is what makes the schedule operationally robust to this non-monotonicity: the working trajectory $\{1, 2, 9\}$ emerges from a nominally nine-stage schedule without manual intervention. In the present strong-form forward training the graded homotopy provides no benefit; the direct elastic-to-full-cracked jump is what trains the network.

A natural follow-up question is whether the intermediate-α non-convergence is a property of the constitutive map or of the training direction: ramping α upward from elastic asks the optimiser to leave a converged basin that has no diagonal-strut topology, so perhaps starting from a converged cracked field and ramping α downward through the same intermediate values would be stable. We tested this directly: warm-starting from a converged stage-8 checkpoint ($\alpha = 1.00,\ S_{\max} = 5$ mm) from a preliminary run that completed the direct jump, and re-attempting each intermediate stage in descending α order with the same ReLoBraLo weighter, fixed arc weight, gradient clip and 2500-iteration budget. The outcomes are reported in Table 6: four of the five intermediate stages again failed to converge, and only the ($\alpha = 0.50,\ S_{\max} = 2$ mm) stage converged.

Table 6
Backward-init backfill test: the five intermediate-α stages of Table 3 re-attempted in descending α order, each warm-started from the prior step's last good weights. Columns give the backfill step, the stage α and $S_{\max}$, the warm-start source and the convergence outcome.

backfill step	α	$S_{\max}$ (mm)	warm-start	outcome
1	0.75	5.0	stage-8 ckpt ($\alpha = 1.00$)	did not converge
2	0.50	5.0	step 1 fallback	did not converge
3	0.50	2.0	step 2 fallback	converged
4	0.25	2.0	step 3 ckpt	did not converge
5	0.10	2.0	step 4 fallback	did not converge

Taken with the forward-direction outcomes of Table 3, this is consistent with the loss-landscape interpretation above: it is the $\alpha \in (0, 1)$ constitutive blend itself, not the curriculum direction, that the optimiser cannot drive to convergence. The simpler three-stage trajectory $\{1, 2, 9\}$ is the right strategy: hopping the elastic and full-cracked basins directly, on the checkpoint-rollback safety net, is operationally robust precisely because it never asks the optimiser to converge to a basin the network never sees in the final problem.

6.3. Arc-length parametrisation versus load parametrisation

The trained network's clear sign change of $d\lambda/ds$ at $s \approx 0.83$ and the subsequent descending branch through to $\lambda = 2.53$ at $\delta = 10$ mm, an independently solved segment of about 1.7 mm, are the operational test of the arc-length parametrisation. A load-controlled forward PINN, by contrast, cannot represent any pair $(\mathbf{u}, \lambda)$ past the limit point because the load factor λ is the network input and cannot decrease while δ continues to grow; the production CSFM solver of Kaufmann et al. [3] and the commercial 3D implementation of IDEA StatiCa [6] both halt at the limit point for the same reason. The arc-length parametrisation removes the singularity of the load-controlled formulation by treating $\lambda(s)$ as an output rather than an input, and the present training campaign demonstrates a forward PINN that folds through the limit point onto an independently solved post-peak segment of a cracked-concrete D-region in a single forward pass.

The trained network and the displacement-controlled reference locate the peak at essentially the same deflection (8.3 vs 8.5 mm); the residual difference is the peak *value*, which the network overshoots by 14%. This is the distributed-versus-localised-branch gap of Section 4, not a peak-location mismatch: the network and the secant reference, both of which use a smooth non-localising field, sit on the upper distributed branch, while a tightly-equilibrated consistent-tangent solver localises the softening into a single element row and reaches a lower branch. The two branches reflect the classical non-uniqueness of unregularised local strain-softening [28, 29]; a well-posed bearing model together with a fracture-energy compressive law (Section 6.4) collapses them to a single mesh-objective curve and is the route to closing the 14% gap.

6.4. Scope and limitations

The trained network is specific to one deep-beam geometry, one reinforcement layout, and one material set. The natural extension is the parametric architecture that consumes a design vector $\boldsymbol{\theta}_{\text{design}}$ and trains on a Latin-hypercube sweep of a populated design family, as in the author's earlier neural-network surrogates [31, 32]; the arc-length parametrisation is unchanged by that extension, only the input dimensionality of the trunk grows. The 5×10^5 inference-cost ratio of Section 5.8 becomes operationally meaningful for parametric CSFM design only after the parametric training is in place. Beyond amortised inference, the trained network is a differentiable map from design and load to the equilibrium path, so gradient-based design optimisation and inverse identification are available without the adjoint machinery an incremental solver requires; this differentiability, not single-curve speed, is the methodological reason to prefer the learned formulation.

Even on this single design, the deep-beam result is reproducible across five seeds, but the peak-deflection spread (5.9 ± 3.0 mm, Section 5.4) is a real limitation, and the corbel and wall-pier cases remain single-seed. The validated output is the scalar equilibrium path, not the pointwise internal fields: those are carried only to form the interior equilibrium loss, which the network as trained satisfies loosely, and nothing in the ansatz enforces the benchmark's left-right symmetry, so the spatial fields are asymmetric (a 50–70% mirror discrepancy on the ascending branch) and are not claimed as a prediction; a symmetry-enforcing ansatz must be confined to the ascending branch, since the post-peak field physically bifurcates at the limit point. The un-anchored post-peak branch is non-unique: warm-extending the

converged network and a from-scratch wide-window solve degenerate in the same way, so the independent accuracy comparison is confined to the ≈ 1.7 mm deep-beam window and the corbel and wall-pier full-range curves are reported as anchored fits rather than independent traces. Both the network and the displacement-controlled reference sit on the upper distributed-softening branch and inherit the mesh dependence of local strain-softening; material-softening regularisations alone do not restore objectivity [33], whereas a well-posed bearing model together with a fracture-energy (G_{fc}) compressive law [27, 29] does (first-order mesh convergence, extrapolated $\lambda_\infty \approx 1.20$) and is what would collapse the distributed and localised branches to a single mesh-objective capacity. Finally, the validation is numerical rather than experimental, the hard initial condition $\mathbf{u}(s = 0) = 0$ currently precludes a non-zero axial dead load (so the wall pier is trained on a no-N variant), and the arc-length constraint is enforced on the loaded patch rather than as Crisfield's global configuration-space norm [9]; a controlled ablation (repository) confirms that the soft λ-floor used on the asymmetric benchmarks acts as a basin selector set from a cheap elastic-FE estimate, not as a fit to the validation target.

7. Conclusions

1. This study presents an arc-length-parametrised physics-informed neural network that folds the equilibrium path of a cracked CSFM discontinuity region from the elastic regime, through the limit point and onto an independently solved post-peak segment, in a single forward pass of the trained network. The parametrisation by an arc-length coordinate s rather than by the load factor λ removes the singularity that halts load-controlled solvers at the limit point.

2. Four methodological ingredients are jointly necessary for the training to converge in the cracked regime: a supervised elastic-FE pre-training that escapes the trivial-zero attractor of the physics losses; a pointwise mean-of-squared-deviations form of the arc-length loss that prevents the Cauchy-Schwarz gaming of the natural squared-deviation-of-mean form; a fixed weight on that loss held outside the adaptive ReLoBraLo pool so it cannot be silently down-weighted; and a C^1-regularised cracked-membrane constitutive that replaces the hard case distinctions of the cracked rotating compression field by sigmoid and softplus blends, removing the NaN gradients that the hard form produces under forward-PINN training.

3. On a benchmark deep-beam D-region the trained network folds through the limit point onto an independently solved post-peak segment of about 1.7 mm (canonical seed) in a single forward pass, identifying the limit point through a sign change of the numerical derivative $\Delta\lambda/\Delta s$ at $s \approx 0.83$. Across five random seeds the peak load factor is $\lambda_{\text{peak}} = 2.64 \pm 0.29$, the peak deflection 5.9 ± 3.0 mm, and the path-RMSE $26 \pm 6\%$ of the reference peak; on the canonical seed (fixed *a priori*) the peak deflection matches the reference (8.3 vs 8.5 mm), the RMSE is 18%, and the peak load factor overshoots by 14%.

4. The seed-robust claim is the qualitative one (that the arc-length parametrisation reproducibly folds through the limit point and begins descending onto the post-peak branch) rather than a tight quantitative bound on the peak location, which the fourfold deflection spread across seeds (Section 5.4) shows is not yet achieved. This independently-resolved post-peak segment lies on the *distributed-softening* branch of the local-secant CSFM idealisation: an independent consistent-tangent path-follower that reaches tight equilibrium localises onto a lower, mesh-dependent branch ($\lambda \approx 1.0$–1.4). The result should therefore be read as a feasibility demonstration that the network independently reproduces the onset of the reference idealisation's post-peak branch over a limited window past the limit point, not as an independent trace of the complete descending branch nor as a prediction of the regularised structural capacity, which a crack-band or nonlocal constitutive would be required to define (Section 6.4).

5. Beyond the symmetric deep beam, the formulation extends to an asymmetric cantilever corbel and a slender wall pier, which require additional boundary-condition-aware ingredients (a directional arc-length constraint and a soft λ-floor for the corbel, and per-iteration gradient sanitisation for the wall pier), establishing that the loss formulation generalises across D-region geometries. For these two benchmarks the full-range curves are anchored, regressed onto the reference rather than solved as independent post-peak traces, so the deep-beam result above remains the study's one independent post-peak demonstration.

6. The trained network evaluates a 200-point arc-length sweep in 0.3 ms on a single CPU core, against 141 s for the displacement-controlled reference. Reported as a forward-pass measurement only: the per-curve cost of training (pre-training plus staged Adam) dominates, so for a single curve the PINN is slower than the reference. The cost

ratio becomes operationally meaningful only under a parametric network covering a design family, which is not built in this study and is the natural follow-up (Section 6.4).

Across the three studies of this programme, the lower-bound capacity, the internal stress field, and the equilibrium path of a structural-concrete D-region are obtained in three forward passes, without the incremental finite-element solve those quantities have traditionally required; the consequential next step is a parametric, experimentally validated extension.

Data and code availability

The training, validation, and reference code for all three benchmarks of this study will be made available upon publication. The differentiable constitutive embedding is shared with the companion inverse-PINN study [19]; the present arc-length parametrisation, the C^1 regularisation, the pretraining remedy, the patch-restricted arc-length loss, the directional, floor, and sanitiser extensions, and the anchor fine-tune step are introduced and validated here.

References

[1] Peter Marti. Basic tools of reinforced concrete beam design. *ACI Journal*, 82(1):46–56, 1985.

[2] Aurelio Muttoni, Joseph Schwartz, and Bruno Thurlimann. *Design of Concrete Structures with Stress Fields*. Birkhauser, 1997.

[3] Walter Kaufmann, Jaime Mata-Falcon, Manuel Weber, and Tena Galkovski. *Compatible Stress Field Design of Structural Concrete*. ETH Zurich, 2020.

[4] Walter Kaufmann and Peter Marti. Structural concrete: Cracked membrane model. *Journal of Structural Engineering*, 124(12):1467–1475, 1998.

[5] Jorg Schlaich, Kurt Schafer, and Mattias Jennewein. Toward a consistent design of structural concrete. *PCI Journal*, 32(3):74–150, 1987.

[6] IDEA StatiCa. IDEA StatiCa Detail: Structural design of concrete 3D discontinuities. Software documentation, 2024. URL https://www.ideastatica.com/support-center/idea-statica-detail-structural-design-of-concrete-3d-discontinuities. Accessed 2026-05-22.

[7] E. Riks. An incremental approach to the solution of snapping and buckling problems. *International Journal of Solids and Structures*, 15(7): 529–551, 1979.

[8] G. A. Wempner. Discrete approximations related to nonlinear theories of solids. *International Journal of Solids and Structures*, 7(11): 1581–1599, 1971.

[9] M. A. Crisfield. A fast incremental/iterative solution procedure that handles "snap-through". *Computers & Structures*, 13(1–3):55–62, 1981.

[10] M. J. Clarke and G. J. Hancock. A study of incremental-iterative strategies for non-linear analyses. *International Journal for Numerical Methods in Engineering*, 29(7):1365–1391, 1990.

[11] Michael H. Scott. Arc length parameters. OpenSees Digital, blog post, 2 May 2023, 2023. URL https://openseesdigital.com/2023/05/02/arc-length-parameters/. Accessed 2026-05-22.

[12] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[13] 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.

[14] Ehsan Haghighat, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379:113741, 2021.

[15] Shahed Rezaei, Ali Harandi, Ahmad Moeineddin, Bai-Xiang Xu, and Stefanie Reese. A mixed formulation for physics-informed neural networks as a potential solver for engineering problems in heterogeneous domains: Comparison with finite element method. *Computer Methods in Applied Mechanics and Engineering*, 401:115616, 2022.

[16] E. Samaniego, C. Anitescu, S. Goswami, V. M. Nguyen-Thanh, H. Guo, K. Hamdia, X. Zhuang, and T. Rabczuk. An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. *Computer Methods in Applied Mechanics and Engineering*, 362:112790, 2020.

[17] Vien Minh Nguyen-Thanh, Xiaoying Zhuang, and Timon Rabczuk. A deep energy method for finite deformation hyperelasticity. *European Journal of Mechanics - A/Solids*, 80:103874, 2020.

[18] Vera M. Balmer, Walter Kaufmann, and Michael A. Kraus. Physics-informed neural networks for nonlinear analysis of reinforced concrete beams. In *Lecture Notes in Civil Engineering*, volume 494. Springer, 2024. doi:10.1007/978-3-031-60271-9_24.

[19] Sandesh Lamsal and Rubi Bhandari. Inverse physics-informed neural network for csfm stress field reconstruction in concrete d-regions. engrXiv preprint: https://doi.org/10.31224/7284, 2026.

[20] Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021. doi: 10.1137/20M1318043.

[21] Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022. doi: 10.1016/j.jcp.2021.110768.

[22] Rafael Bischof and Michael A. Kraus. Multi-objective loss balancing for physics-informed deep learning. *Computer Methods in Applied Mechanics and Engineering*, 439:117914, 2025. doi: 10.1016/j.cma.2025.117914.

[23] Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems*, 34:26548–26560, 2021.

[24] European Committee for Standardization (CEN). *Eurocode 2: Design of Concrete Structures — Part 1-1: General Rules and Rules for Buildings (EN 1992-1-1)*. European Committee for Standardization (CEN), Brussels, 2004.

[25] Frank J. Vecchio and Michael P. Collins. The modified compression-field theory for reinforced concrete elements subjected to shear. *ACI Journal*, 83(2):219–231, 1986.

[26] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. *arXiv preprint*, 2017. arXiv:1710.05941.

[27] Zdeněk P. Bažant and Byung H. Oh. Crack band theory for fracture of concrete. *Matériaux et Constructions*, 16(93):155–177, 1983.

[28] Zdeněk P. Bažant and Feng-Bao Lin. Nonlocal smeared cracking model for concrete fracture. *Journal of Structural Engineering*, 114(11): 2493–2510, 1988.

[29] Zdeněk P. Bažant. Concrete fracture models: Testing and practice. *Engineering Fracture Mechanics*, 69(2):165–205, 2002.

[30] Martin Bimschas. *Displacement-Based Seismic Assessment of Existing Bridges in Regions of Moderate Seismicity*. PhD thesis, ETH Zurich, 2010. Doctoral dissertation, ETH Zurich, Switzerland.

[31] Sandesh Lamsal and Rubi Bhandari. A strut and tie neural network surrogate for failure-load prediction of concrete d-regions designed by the compatible stress field method. engrXiv preprint: https://doi.org/10.31224/7250, 2026.

[32] Sandesh Lamsal. DeepForm: A graph neural network surrogate for real-time tensile membrane form-finding across anticlastic typologies. engrXiv preprint: https://doi.org/10.31224/7269, 2026.

[33] Gilles Pijaudier-Cabot and Zdeněk P. Bažant. Nonlocal damage theory. *Journal of Engineering Mechanics*, 113(10):1512–1533, 1987.