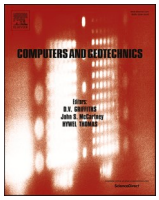
Computers and Geotechnics 188 (2025) 107587

Contents lists available at ScienceDirect

Computers and Geotechnics

journal homepage: www.elsevier.com/locate/compgeo



Material characteristic length insensitive nonlocal modelling: A computationally efficient scaled nonlocal integral method

Xin Zhou [a,b], Dechun Lu [a,*], Jidong Zhao [b], Yaning Zhang [a], Zhiwei Gao [a,c], Timon Rabczuk [a,d], Xiuli Du [a]

[a] Institute of Geotechnical and Underground Engineering, Beijing University of Technology, Beijing 100124, China
[b] Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong 999077, China
[c] James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK
[d] Institute of Structural Mechanics, Bauhaus-Universität Weimar, Weimar 99423, Germany

ARTICLE INFO

Keywords:
Nonlocal model
Characteristic length
Stress update algorithm
Damage
Plasticity

ABSTRACT

Nonlocal modelling has achieved notable progress in resolving mesh dependence but remains constrained by two persistent challenges: sensitivity to characteristic length parameters and high computational costs. This study presents a scaled nonlocal integral formulation coupled with an optimized computational framework to simultaneously address two limitations. We first analytically demonstrate that variations in characteristic length induce proportional scaling of load–displacement curves, revealing that apparent changes in structure softening rate are artifacts of this scaling. Building on this insight, a dimensionless scaling factor is derived to systematically eliminate characteristic length dependence, enabling consistent predictions across parameter choices. The proposed method is integrated with a Mohr-Coulomb plasticity damage model, employing a return mapping algorithm for plasticity and a novel hybrid local-nonlocal solver accelerated by octree spatial partitioning for damage evolution. Three benchmark boundary value problems, evaluated across diverse element sizes, characteristic lengths, and softening laws, validate the robustness of the method. The results demonstrate that the proposed nonlocal method achieves mesh- and length-invariant load–displacement responses while accommodating arbitrary softening functions. The presented nonlocal computation method also shows a remarkable computational efficiency compared to the traditional nonlocal computation method.

1. Introduction

Numerical simulation of material failure due to strain softening has long been a major challenge in the field of computational solid mechanics. Local continuum models encounter difficulties in simulating material failure because of the problem of ill-conditioned mesh dependence. This issue primarily manifests in the fact that, with mesh refinement, the slope of the softening segment in the load–displacement response becomes steeper, the localization zone narrows infinitely, and the energy dissipated within this zone approaches zero. The localization zone is identified as the shear band in ductile materials or the fracture process zone in brittle materials (Wu et al., 2024; Yu et al., 2024). Mathematically, this phenomenon can be explained by the loss of ellipticity in governing equations during the strain softening (Anand et al., 2012; Mazzucco et al., 2019), leading to pathological boundary value problems where numerical solutions depend on element size.

Several nonlocal methods have been proposed in the literature to address this limitation, including the gradient-dependent method (Engelen et al., 2003; Zhao et al., 2005), the Cosserat method (de Borst, 1991), and the nonlocal integral method (Bažant and Jirásek, 2002; Lü et al., 2023). These nonlocal methods assume that the response of a material point is influenced by both its own state and the state of neighbourhood points (Cui et al., 2023; Ren et al., 2020; Yu and Sun, 2021). The degree of strain localization is controlled by a scale parameter known as the characteristic length (Karapiperis et al., 2021). Among the various nonlocal methods, the phase field method (Miehe et al., 2010) and Peridynamics (Silling and Lehoucq, 2010) have garnered significant attention from researchers in recent years (Yang et al., 2024). The phase field method, sometimes regarded as a special gradient damage method (de Borst and Verhoosel, 2016), employs a continuous phase field variable to describe sharp crack topology. The evolution of the phase field variable is based on minimizing the system's total energy. Therefore, crack propagation is automatically tracked, eliminating

* Corresponding author.
E-mail address: dechun@bjut.edu.cn (D. Lu).

https://doi.org/10.1016/j.compgeo.2025.107587
Received 6 June 2025; Received in revised form 19 August 2025; Accepted 20 August 2025
Available online 30 August 2025

Nomenclature

Notation

$\boldsymbol{\sigma}$	nominal stress tensor
$\overline{\boldsymbol{\sigma}}$,$\overline{\boldsymbol{s}}$	effective stress tensor, effective deviatoric stress tensor
J_2,J_3	second and third invariants of $\overline{\boldsymbol{\sigma}}$
$\overline{p}$, $\overline{q}$,$\overline{\theta}$	mean stress, generalized shear stress, and Lode angle of $\overline{\boldsymbol{\sigma}}$
$\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}^{\mathrm{e}}$,$\boldsymbol{\varepsilon}^{\mathrm{p}}$	total, elastic, and plastic strain tensors
$\varepsilon_{\mathrm{d}}^{\mathrm{p}}$	equivalent plastic shear strain
$\mathrm{d}\phi$,κ^{p}	plastic multiplier, plastic internal variable
κ^{d},$\widehat{\kappa}^{\mathrm{d}}$	local and nonlocal damage internal variables
ω	damage variable
H	softening function
A_s	softening parameter
$\mathbb{D}_0$	undamaged elastic stiffness matrix
E	Young's modulus
l_c,R	characteristic length, neighbourhood radius
l_{FE}	element size
m	interpolation coefficient in the over-nonlocal formulation
w	width of the localization zone
w_∞	weight function
η	scaling factor
f,g	yield function, plastic potential function
R_{mc}	shape function of f on the meridian plane
R_{mw}	shape function of g on the meridian plane
c_0, φ,ψ	initial cohesive strength, internal friction angle, dilation angle
β	smooth parameter in the R_{mc}
ϵ	meridional eccentricity
Φ	merit function
ρ,ζ	algorithm parameters in the line search method

the need for tedious crack-tracking tasks (Wu, 2017). For instance, Chen et al. (2025) developed a high-accuracy phase field model that can accurately reproduce arbitrary cohesive laws. Furthermore, Feng and Hai (2025) proposed a unified directional energy decomposition method to incorporate cohesive laws into the phase field model in a 3D setting. Peridynamics uses governing integral equations instead of partial differential equations (PDEs) in continuum mechanics (Dorduncu et al., 2024), avoiding the singularity caused by the absence of spatial derivatives at discontinuities such as cracks. It describes material damage and cracking through the breakage of "bonds", making it highly suitable for solving discontinuous deformation and failure problems.

Generally, the nonlocal models have successfully regularized the ill-posed mesh dependence problem; however, they also present the following challenges: (*i*) the simulation results depend on the characteristic length; (*ii*) the numerical implementation incurs high computational costs. This paper focuses on the nonlocal integral method, as it does not require modification to governing PDEs compared to the local model, is more easily implemented numerically, and has received considerable attention in recent research (Li and Gao, 2024; Su et al., 2023). It is well-recognized that the characteristic length significantly influences the global response of a structure after strain localization. Generally, as the characteristic length increases, the localization zone width predicted by the nonlocal model increases (Mallikarachchi and Soga, 2020), while the softening rate of the load–displacement response decreases. Typically, two approaches exist for determining the characteristic length. The first scheme treats the characteristic length as a material parameter that influences the load–displacement response. The characteristic length is determined through inverse calibration based on structural results, as demonstrated by Bažant and Pijaudier-Cabot (1989), Le Bellégo et al. (2003), Jirásek et al. (2004), and Iacono et al. (2006). However, as noted by Xenos et al. (2015), an accurate alignment with structural findings may result in an excessively exaggerated localization zone width. For example, by fitting experimental data on nominal fracture energy and nominal strength of concrete, Jirásek et al. (2004) obtained a nonlocal radius of approximately 75 mm, which significantly exceeded the one observed in experiments. The localization zone width measured in experiments is about 2.5–4 times the maximum aggregate size $d_{\max}$ for concrete materials (Denarie et al., 2001; Otsuka and Date, 2000) and about 8–20 times the mean diameter d_{50} for granular materials (Mohamed and Gutierrez, 2010; Rattez et al., 2022). It is often not easier to adjust the characteristic length by matching it to the localization zone observed in the experiment. This is because the nonlocal model is generally valid only when the characteristic length exceeds the element size, meaning that accurate modelling of the localization zone requires extremely fine meshes, making the cost of numerical analysis for large-scale engineering problems extremely high.

Another approach is to treat the characteristic length as a numerical parameter that describes the localization zone width. The influence of characteristic length on the load–displacement response is attempted to be eliminated by appropriately adjusting the material parameters that control the softening rate. This method, called the softening modulus modification method, was introduced by Marcher (2003) and has since been widely adopted in subsequent nonlocal modelling, such as by Galavi and Schweiger (2010), Mánica et al. (2018), Singh et al. (2021), and Shi et al. (2023). Numerical simulation results from these studies suggest that a unique load–displacement response may be obtained by fixing the ratio of model parameters controlling the softening rate to the characteristic length. However, strictly speaking, the scaling softening rate method is a semi-empirical approach based on numerical simulation laws. It has not been proven to be valid for all types of softening functions. Indeed, the impact of model parameters on softening behaviour depends on the specific form of the softening function (Sun et al., 2021). In some constitutive models, multiple parameters may even influence the softening rate (Zhou et al., 2020). It is imperative to recognize that eliminating the characteristic length sensitivity requires a fundamental understanding of how characteristic length affects the global response. This is also the primary research motivation of this paper. In this study, we will demonstrate analytically that the load–displacement curve is substantially scaled by the change in characteristic length, and changes in structural softening rate are merely a superficial phenomenon. Based on this fundamental understanding, we propose a novel nonlocal method that is insensitive to material characteristic length by scaling the nonlocal variable instead of the softening parameter, making it valid for any softening function.

High computational costs pose another challenge that hinders the application of nonlocal models in practical engineering problems. This issue arises from the computational rules governing the nonlocal variables(s). In numerical computation, the relevant local variable at each integration point in the vicinity of the present integration point is weight averaged to obtain the nonlocal variable. This necessitates initially identifying neighbourhood integration points and calculating the corresponding weight coefficients for nonlocal computations. Numerous efforts have been made to enhance nonlocal models' computational efficiency. Jirásek (2007) proposed initially computing the weight coefficients and storing them for subsequent calculations. However, this approach overlooks potential changes in the number and position of neighbourhood integration points during the calculation process. Chen et al. (2023) suggested performing nonlocal calculations every T increment steps. As T increases, the nonlocal computational cost decreases. Selecting an appropriate value for T is essential to balance

computational accuracy and efficiency. Moreau et al. (2015) proposed a nonlocal computing scheme that integrates a thick-level set approach, which limits nonlocal computation to a scope marginally greater than the shear band. In the algorithm proposed by Tang et al. (2021), the search scope is limited to a smaller region, as opposed to considering all integration points. Specifically, it focuses on the neighbourhood integration points of the target integration point, along with those of neighbourhood integration points from the preceding time step. The core idea of the above computing schemes is to reduce the scope of nonlocal calculations spatially or to reduce the frequency of nonlocal calculations temporally. Alternatively, the computational efficiency of nonlocal models can be further improved by combining efficient neighbourhood search algorithms such as the r-tree algorithm, k-d tree algorithm, or the octree algorithm. Compared to the widely used traversal algorithm in nonlocal computing, these algorithms theoretically have lower time complexity and are more suitable for performing large-scale data search tasks. For instance, Fernández-Fernández et al. (2022) improved the neighbourhood search efficiency of SPH (Smoothed Particle Hydrodynamics) simulation by employing the octree algorithm. Compared to the well-established Cell-Linked-List method, their method achieves up to a 1.9 × speedup in the numerical simulations involving millions of particles. Through numerical analyses involving varying numbers of material points, Vazic et al. (2020) evaluated the performance of several family search algorithms in Peridynamics simulations, including traverse search, region partitioning, enhanced r-tree with packing, and balanced k-d tree algorithms. The results indicate that traverse search exhibits the poorest performance, whereas the boosted r-tree algorithm emerges as the most efficient among the four evaluated. This paper presents an efficient nonlocal computational framework that combines the octree algorithm with a local-nonlocal hybrid computing strategy to enhance nonlocal models' computational efficiency in large-scale numerical analyses.

The remainder of this paper is organized as follows: In Section 2, we first briefly review the fundamentals of the nonlocal integral method and then reveal the influence mechanism of characteristic length on the load–displacement response from an analytical perspective. A scaled nonlocal method is proposed to eliminate the characteristic length sensitivity of the nonlocal model. In Section 3, we apply the proposed method to a Mohr-Coulomb plasticity damage model. Two acceleration strategies are employed in the numerical implementation of the model. In Section 4, numerical examples with varying element sizes and characteristic lengths are used to evaluate the effectiveness of the proposed methods. The main conclusions are presented in Section 5.

2. Scaled nonlocal integral method

The fundamental concepts of the nonlocal method are briefly reviewed. Building upon the over-nonlocal formulation (Di Luzio and Bažant, 2005), an analytical solution is derived for a 1D boundary value problem, elucidating the intrinsic relationship between characteristic length and load–displacement response. Finally, a scaled nonlocal method, insensitive to the material characteristic length, is proposed.

2.1. Nonlocal integral method

The over-nonlocal formulation is defined as the linear interpolation between the standard nonlocal constitutive variable and the local constitutive variable. Previous studies (Huang et al., 2018; Lu et al., 2009) have demonstrated that the over-nonlocal formulation effectively preserves the PDEs' well-posedness when using the strain-softening model. The over-nonlocal formulation is expressed as:

$$\widehat{\kappa}(\boldsymbol{x}) = (1-m)\kappa(\boldsymbol{x}) + m\frac{\int_V w_\infty(\boldsymbol{x},\boldsymbol{\xi})\kappa(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}}{\int_V w_\infty(\boldsymbol{x},\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}} \tag{1}$$

where κ and $\widehat{\kappa}$ respectively represent the local variable and the corresponding nonlocal variable. m is an interpolation coefficient. Eq. (1) degenerates into a local formula when $m = 0$ and into a standard nonlocal formula when $m = 1$. V denotes the integral domain. $w_\infty(\boldsymbol{x},\boldsymbol{\xi})$ is a spatial weight function centred around the coordinate $\boldsymbol{x}$. In this work, the weight function of the bilinear exponential type shown in Fig. 1(a) is adopted, defined as follows:

$$w_\infty(\boldsymbol{x},\boldsymbol{\xi}) = \frac{1}{a_w}\mathrm{e}^{-\frac{\|\boldsymbol{\xi}-\boldsymbol{x}\|}{l_c}} \tag{2}$$

where l_c is the characteristic length. a_w is a constant used to normalize the weight function, $a_w = 2l_c$ for the 1D conditions, $a_w = 2\pi l_c^2$ for the 2D conditions, and $a_w = 8\pi l_c^3$ for the 3D conditions.

Generally, the weight function will adopt a truncated form to enhance the computational efficiency of the nonlocal model.

$$w_\infty(\boldsymbol{x},\boldsymbol{\xi}) = \begin{cases} w_\infty, & \|\boldsymbol{\xi}-\boldsymbol{x}\| \le R \\ 0, & \|\boldsymbol{\xi}-\boldsymbol{x}\| > R \end{cases} \tag{3}$$

where R denotes the neighbourhood radius, as shown in Fig. 1(b). A default value of $R = 2l_c$ (Galavi and Schweiger, 2010) is adopted in this paper.

2.2. Localization analysis and scaled nonlocal method

To explore the essential influence of characteristic length on the global response, we first narrow the research scope to the 1D condition to simplify the analysis. As depicted in Fig. 2, the bar has a fixed left end and a right end that is subjected to a tensile load. The length of the bar is denoted as L, and the displacement at the right end is represented by u. Without loss of generality, the strain localization is assumed to start at the neutral axis $x = 0$ of the bar. When the stress reaches the yield stress, the plastic deformation occurs within a localization zone of width w, while elastic unloading occurs outside the localization zone.

During the plastic deformation stage, the governing equations of the 1D local elastoplasticity model is expressed as:

$$\sigma = E(\varepsilon - \varepsilon^{\mathrm{p}}) \tag{4}$$

$$f(\sigma,\varepsilon^{\mathrm{p}}) = \sigma - \sigma_0 H(\varepsilon^{\mathrm{p}}) = 0 \tag{5}$$

where σ, ε, and ε^{p} represent the stress, total strain, and plastic strain, respectively. $H(\cdot)$ represents the normalized softening function. E represents Young's modulus. σ_0 denotes the yield stress.

The key to developing a nonlocal model from a local model lies in replacing the constitutive variable controlling the strain-softening behaviour with the nonlocal variable. The nonlocal plasticity model corresponding to Eq. (4) and Eq. (5) is represented by:

$$\sigma = E(\varepsilon - \widehat{\varepsilon}^{\mathrm{p}}) \tag{6}$$

$$f(\sigma,\widehat{\varepsilon}^{\mathrm{p}}) = \sigma - \sigma_0 H(\widehat{\varepsilon}^{\mathrm{p}}) = 0 \tag{7}$$

where $\widehat{\varepsilon}^{\mathrm{p}}$ is the nonlocal plastic strain.

Eq. (7) can be written as:

$$\widehat{\varepsilon}^{\mathrm{p}}(x) = H^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{8}$$

where $H^{-1}(\cdot)$ represents the inverse function of $H(\cdot)$.

Substituting Eq. (1) and Eq. (2) into Eq. (8), we can obtain:

$$\varepsilon^{\mathrm{p}}(x) + \frac{m}{2l_c(1-m)}\int_{-w/2}^{w/2}\mathrm{e}^{\frac{-|\xi-x|}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi = \frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{9}$$

The form of Eq. (9) satisfies the Fredholm integral equation of the second kind (Polyanin and Manzhirov, 2008). According to the derivation in Appendix A, the solution of $\varepsilon^{\mathrm{p}}(x)$ within the strain localization zone is given with the limitation condition of $m > 1$ as follows:

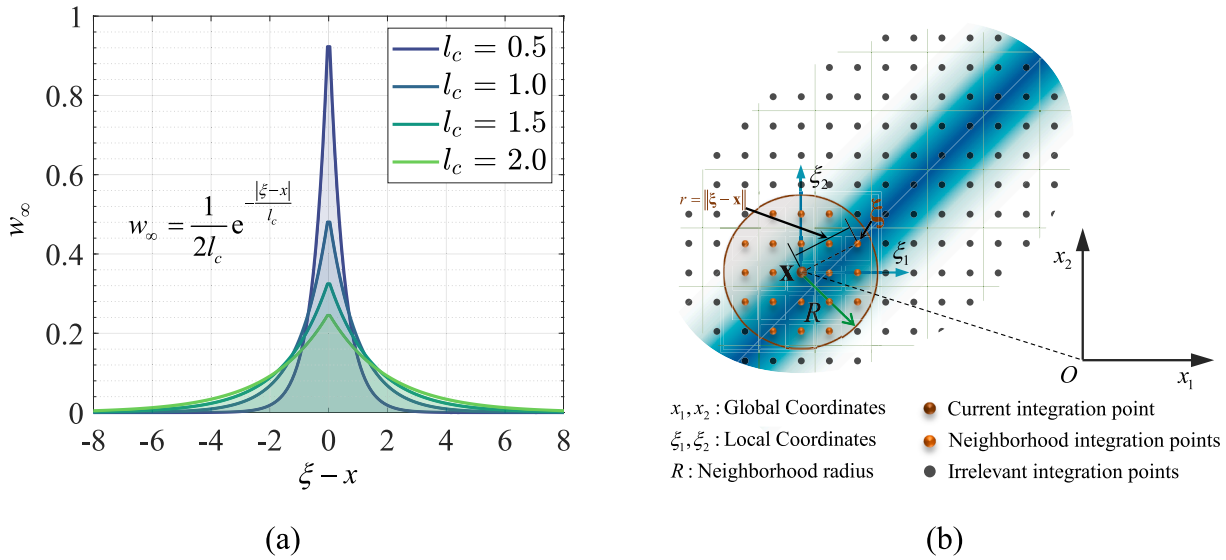


Fig. 1. Nonlocal integral method: (a) weight function curves; (b) integration points and interaction domain.

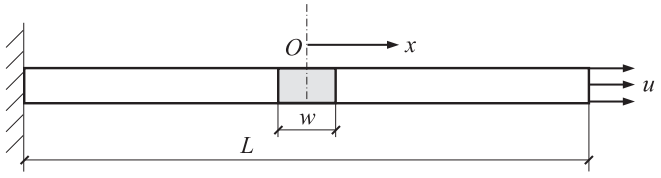


Fig. 2. Geometry and boundary conditions of the bar.

$$\varepsilon^{\mathrm{p}}(x) = H^{-1}\left(\frac{\sigma}{\sigma_0}\right)\left[\sqrt{\frac{m}{m-1}}\cos\left(\frac{x}{l_c\sqrt{m-1}}\right)+1\right],\quad -\frac{w}{2}\le x\le\frac{w}{2} \tag{10}$$

where the expression for the localization zone width w can be found in Eq. (A39).

By integrating Eq. (10) over the interval $[-w/2 \le x \le w/2]$, the displacement u^{p} of the right end of the bar induced by the plastic strain is obtained by:

$$u^{\mathrm{p}} = wH^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{11}$$

In what follows, we first present the relationship between load (F) and plastic displacement (u^{p}), as the characteristic length primarily affects the plastic deformation behaviour of the bar. Eq. (11) can be easily rewritten as:

$$\sigma = \sigma_0 H\left(\frac{u^{\mathrm{p}}}{w}\right) \tag{12}$$

Substituting the expression of w presented in Eq. (A39) into Eq. (12) and multiplying both sides by the cross-sectional area S of the bar, one can obtain the relationship between the load and the plastic displacement as follows:

$$F = \sigma S = \sigma_0 SH\left(\frac{1}{d_m}\frac{u^{\mathrm{p}}}{l_c}\right) \tag{13}$$

where $d_m = 2\sqrt{m-1}\left(\pi - \arctan\left(1/\sqrt{m-1}\right)\right)$. From Eq. (A39), we know that $w = d_m l_c$. m is required to be greater than 1 for the localization analysis. Details can be found in Appendix A.

From Eq. (13), it is evident that l_c appears in the denominator, implying that modifying l_c essentially scales the load–displacement response on the plastic displacement axis. The apparent changes in structure softening rate are artifacts of this scaling, as shown in Fig. 3.

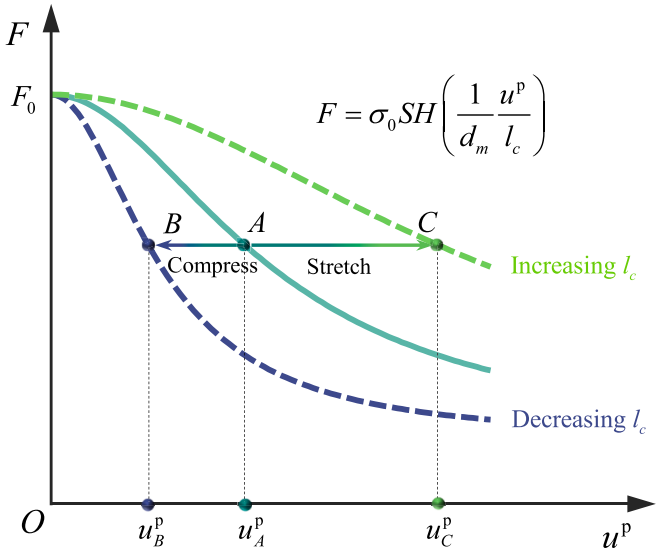


Fig. 3. Influence of characteristic length change on the load–displacement curve.

Based on this insight, let's assume that the influence of l_c on load displacement response can be counteracted by introducing a dimensionless scaling factor η into the softening function of Eq. (7).

$$\sigma - \sigma_0 H\left(\eta\hat{\varepsilon}_m^{\mathrm{p}}\right) = 0 \tag{14}$$

By using Eq. (14) instead of Eq. (7), Eq. (9) can be re-written as follows:

$$\varepsilon^{\mathrm{p}}(x) + \frac{m}{2l_c(1-m)}\int_{-w/2}^{w/2} e^{\frac{-|\xi-x|}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi = \frac{1}{\eta}\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{15}$$

Repeating the solving process of the Fredholm integral equation in Appendix A and Appendix C, we can yield:

$$\varepsilon^{\mathrm{p}}(x) = \frac{H^{-1}(\sigma/\sigma_0)}{\eta}\left[\sqrt{\frac{m}{m-1}}\cos\left(\frac{x}{l_c\sqrt{m-1}}\right)+1\right],\quad -\frac{w}{2}\le x\le\frac{w}{2} \tag{16}$$

It is worth emphasizing that in the derivation process of Eq. (16), $1/\eta$ can be regarded as a constant multiplied by $H^{-1}(\sigma/\sigma_0)$, which does not affect the solution of the shear band width w. By integrating Eq. (16)

over the interval $[-w/2 \le x \le w/2]$, we can yield:

$$u^{\mathrm{p}} = \frac{w}{\eta} H^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{17}$$

By analogy with the derivation process of Eq. (11) to Eq. (13), the analytical relationship between load and plastic displacement can be obtained again as follows:

$$F = \sigma_0 S H\left(\frac{1}{d_m}\frac{\eta}{l_c}u^{\mathrm{p}}\right) \tag{18}$$

When l_c changes to l_c^s, η is required to be adjusted to η^s to ensure the unique load–displacement response during the softening stage. Then, Eq. (18) can be rewritten as:

$$F = \sigma_0 S H\left(\frac{1}{d_m}\frac{\eta^s}{l_c^s}u^{\mathrm{p}}\right) \tag{19}$$

It's not hard to see that the scaling factor and characteristic length satisfy the following constraint equation:

$$\frac{\eta}{l_c} = \frac{\eta^s}{l_c^s} \tag{20}$$

Eq. (20) is the core formula of the proposed scaled nonlocal method, which applies a scaling factor to adjust the nonlocal variable, compensating for the influences of characteristic length variations on the load–displacement response. The proposed method is valid for any softening law, as the derivation process does not assume a specific form of the softening function. It is also simple, requiring only the multiplication of a scaling factor with the nonlocal variable, without any other adjustments. When the characteristic length changes, the scaling factor is adjusted proportionally. In practical applications, we can simply make $l_c/l_{FE} = \text{constant} > 1$ to obtain the load–displacement responses insensitive to both element size (l_{FE}) and characteristic length. See the following Section 4.3 for details.

In what follows, the analytical relationship between the load (F) and the total displacement (u) is derived. The total strain of the bar can be expressed as follows:

$$\varepsilon(x) = \begin{cases} \varepsilon^{\mathrm{p}}(x) + \dfrac{\sigma}{E} & -\dfrac{w}{2} \le x \le \dfrac{w}{2} \\ \dfrac{\sigma}{E} & -\dfrac{L}{2} \le x < -\dfrac{w}{2},\ \dfrac{w}{2} < x \le \dfrac{L}{2} \end{cases} \tag{21}$$

Integrating Eq. (21) over the interval $[-L/2 \le x \le L/2]$ yields:

$$u = u^{\mathrm{p}} + \frac{\sigma L}{E} = \frac{w}{\eta} H^{-1}\left(\frac{\sigma}{\sigma_0}\right) + \frac{\sigma L}{E} \tag{22}$$

It is difficult to derive the analytic $F(u)$ without giving the specific form of $H(\cdot)$. Let's assume $H(\widehat{\varepsilon}^{\mathrm{p}}) = 1 - A_s\widehat{\varepsilon}^{\mathrm{p}}$, where A_s is the softening parameter. Then, Eq. (22) can be expressed as follows:

$$u = \frac{w}{\eta}\frac{(\sigma_0 - \sigma)}{A_s\sigma_0} + \frac{\sigma L}{E} \tag{23}$$

Performing some manipulations yields:

$$\sigma = \frac{u - \frac{w}{\eta A_s}}{\frac{L}{E} - \frac{w}{\eta A_s \sigma_0}} \tag{24}$$

Substituting Eq. (A39) into Eq. (24) and multiplying both sides by S, we can obtain the relationship between the load and the total displacement as follows:

$$F = \frac{uS - \frac{Sd_m}{A_s}\frac{l_c}{\eta}}{\frac{L}{E} - \frac{d_m}{A_s\sigma_0}\frac{l_c}{\eta}} \tag{25}$$

It can be found that the influence of l_c on the load-total displacement response can still be eliminated by using the scaling formula Eq. (20).

Remark 1. *The softening rate scaling method compensates for the influence of changing l_c on the load–displacement response by adjusting the material parameter that controls the softening rate in the constitutive model. Although this method does not inherently reveal the relationship between l_c and the load–displacement response, they may still be valid in certain cases. When the softening parameter is the coefficient of the nonlocal variable, adjusting the softening coefficient essentially scales the nonlocal variable. But this is just a coincidence. Not all softening functions satisfy this condition.*

Remark 2. *The derivation process presented above assumes a uniform stress distribution along the 1D bar. However, it is important to emphasize that when a stress gradient is present, the size of the localization zone becomes load-dependent and is not solely governed by the material characteristic length* (Challamel et al., 2008a; Challamel et al., 2008b). *In the proposed scaled nonlocal method, the load-total displacement response is not sensitive to the material characteristic length. Instead, it remains dependent on structural dimensions, such as the length scale of the bar, as shown in Eq. (25).*

Remark 3. *In particular, in order to prevent scaled nonlocal method from modifying the model response in the homogeneous case, a judgment condition can be considered by $\eta = 1$ if $|\widehat{\kappa} - \kappa| \le T_e$ where T_e denote a small threshold. Its meaning is that when the structure is in the stage of uniform deformation there is no need to perform the scaling operation because the change in characteristic length will not change the load–displacement response.*

3. Application of scaled nonlocal method in the Mohr-Coulomb plasticity damage model

The proposed scaled nonlocal approach is applied to a Mohr-Coulomb plasticity damage model, which is based on the concept of effective stress which refers to the stress acting on the undamaged configuration of the material. The nominal softening behaviour of the damaged configuration is controlled by the damage variable. The governing equations for the plastic component can be summarized as follows:

$$\begin{cases} \mathrm{d}\overline{\boldsymbol{\sigma}} = \mathbb{D}_0 : (\mathrm{d}\boldsymbol{\varepsilon} - \mathrm{d}\boldsymbol{\varepsilon}^{\mathrm{p}}) & \textit{Hooke's law} \\ \mathrm{d}\boldsymbol{\varepsilon}^{\mathrm{p}} = \mathrm{d}\phi\boldsymbol{r} & \textit{Flow rule} \\ \mathrm{d}\kappa^{\mathrm{p}} = \mathrm{d}\phi h^{\mathrm{p}} & \textit{Hardening law} \\ \mathrm{d}\phi \ge 0, f(\overline{\boldsymbol{\sigma}}, \kappa^{\mathrm{p}}) \le 0, \text{ and } \mathrm{d}\phi f(\overline{\boldsymbol{\sigma}}, \kappa^{\mathrm{p}}) = 0 & \textit{Karush} - \textit{Kuhn} - \textit{Tucker conditions} \end{cases} \tag{26}$$

where $\mathbb{D}_0$ is the undamaged elastic stiffness matrix. $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}^{\mathrm{p}}$ denote the total and plastic strain tensors, respectively. $\overline{\boldsymbol{\sigma}}$ represents the effective stress tensor. $\mathrm{d}\phi$ is the plastic multiplier. In the non-associated flow rule, the plastic flow direction denoted by $\boldsymbol{r}$ is determined by the gradient $\partial g/\partial\overline{\boldsymbol{\sigma}}$ of the plastic potential function g. κ^{p} is the plastic internal variable. h^{p} is the gradient of κ^{p}, and its functional form is related to the selected internal variable form. $f(\overline{\boldsymbol{\sigma}}, \kappa^{\mathrm{p}})$ denotes the yield function.

The nominal stress tensor can be obtained by the strain equivalence hypothesis:

$$\boldsymbol{\sigma} = (1-\omega)\overline{\boldsymbol{\sigma}} = (1-\omega)\mathbb{D}_0 : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}}) \tag{27}$$

where $\boldsymbol{\sigma}$ is the nominal stress tensor acting on the damaged configuration of the material. $\omega(\widehat{\kappa}^{\mathrm{d}})$ and $\widehat{\kappa}^{\mathrm{d}}$ denote the damage variable and the nonlocal damage internal variable, respectively. In the following content, details about the constitutive model are first elucidated. Then, the model implementation is introduced from both plasticity and damage aspects.

3.1. Mohr-coulomb plasticity damage model

In the numerical implementation, the Mohr-Coulomb yield function is typically expressed as a function of the following stress invariants:

$$f = R_{mc}(\overline{\theta})\overline{q} + \overline{M}\overline{p} - \overline{K} \tag{28}$$

where $\overline{q}$, $\overline{p}$, and $\overline{\theta}$ are generalized shear stress, mean stress, and Lode angle defined by the effective stress.

$$\overline{p} = \frac{1}{3}\overline{\boldsymbol{\sigma}} : \mathbf{1} \tag{29}$$

$$\overline{q} = \sqrt{3\overline{J}_2} \tag{30}$$

$$\overline{\theta} = \frac{1}{3}\arccos\left(\frac{3\sqrt{3}\overline{J}_3}{2\overline{J}_2^{3/2}}\right) \tag{31}$$

where $\overline{J}_2 = \frac{1}{2}\overline{\boldsymbol{s}} : \overline{\boldsymbol{s}}$ and $\overline{J}_3 = \det(\overline{\boldsymbol{s}})$ represent the invariants of stress deviator $\overline{\boldsymbol{s}} = \overline{\boldsymbol{\sigma}} - \overline{p}\mathbf{1}$ where $\mathbf{1}$ is the 2nd-order unit tensor. The introduction of $R_{mc}(\overline{\theta})$ can be found in Appendix B.

$\overline{M}$ and $\overline{K}$ are expressed as:

$$\overline{M} = 6\sin\varphi / \left(\sqrt{3}\,(3 - \sin\varphi)\right) \tag{32}$$

$$\overline{K} = 6\overline{c}\cos\varphi / \left(\sqrt{3}\,(3 - \sin\varphi)\right) \tag{33}$$

where φ denotes the internal friction angle. $\overline{c}$ is the effective cohesion, which is considered as the hardening parameter:

$$\overline{c} = c_0 H^{\mathrm{p}}(\kappa^{\mathrm{p}}) \tag{34}$$

where c_0 is the initial cohesive strength. H^{p} is the hardening function. In the presented model, ideal plasticity is adopted, i.e., $\overline{c} = c_0$. The equivalent plastic shear strain $\varepsilon_{\mathrm{d}}^{\mathrm{p}}$ is considered as the plastic internal variable, i.e., $\kappa^{\mathrm{p}} = \varepsilon_{\mathrm{d}}^{\mathrm{p}}$, and thus Eq. (26)$_2$ can be reformulated as:

$$\kappa^{\mathrm{p}} = \mathrm{d}\phi \frac{\partial g}{\partial \overline{q}} \tag{35}$$

where g denotes the plastic potential function. The form of the plastic potential function suggested by ABAQUS software is adopted in this paper.

$$g = \sqrt{(\epsilon c_0)^2 + (R_{mw}\overline{q})^2} + \overline{p}\tan\psi - c_0 \tag{36}$$

where ϵ is the meridional eccentricity. The default value of ϵ is 0.1. ψ is the dilatancy angle. R_{mw} is the shape function of g, the details of which are also introduced in Appendix B.

The following three common forms of damage variable will be considered in the numerical examples of Section 4 to evaluate the effectiveness of the scaled nonlocal method.

$$\omega = \begin{cases} 1 - A_s\widehat{\kappa}^{\mathrm{d}} & \text{Linear type} \\ 1 - \exp\left(A_s\widehat{\kappa}^{\mathrm{d}}\right) & \text{Exponential type} \\ 1 - \dfrac{\widehat{\kappa}^{\mathrm{d}} + 1}{A_s(\kappa^{\mathrm{d}})^2 + \widehat{\kappa}^{\mathrm{d}} + 1} & \text{Rational fraction type} \end{cases} \tag{37}$$

where $\widehat{\kappa}^{\mathrm{d}}$ denotes the nonlocal damage internal variable, i.e., $\widehat{\kappa}^{\mathrm{d}} = \widehat{\varepsilon}_{\mathrm{d}}^{\mathrm{p}}$. In the presented model, the equivalent plastic shear strain is considered as both the damage and plastic internal variables simultaneously. Here, we adopt the suggestion by Grassl and Jirásek (2006) that only $\widehat{\kappa}^{\mathrm{d}}$ is nonlocal while κ^{p} remains local. The advantage of this consideration lies in the consistency condition that needs to be satisfied during the plastic calculation process, which is independent of the nonlocal calculation and helps to improve the stability of numerical calculations.

Under the 1D condition, Eq. (27) can yield:

$$\sigma = (1 - \omega)\overline{\sigma} \tag{38}$$

For ideal plasticity, the effective stress equals the initial yield stress during loading, i.e., $\overline{\sigma} = \sigma_0$. Eq. (38) can be written as:

$$\sigma = (1 - \omega)\sigma_0 \tag{39}$$

Considering Eq. (39) and Eq. (5), it can be found that the damage variable has the following relationship with the softening function:

$$H = 1 - \omega \tag{40}$$

Substituting Eq. (40) into Eq. (39), Eq. (39) can be transformed into Eq. (8), which indicates that the scaled nonlocal method is also applicable to the plasticity damage model.

3.2. Plastic calculation

Using the backward Euler method, Eq. (26) is transformed as:

$$\begin{cases} \overline{\boldsymbol{\sigma}}_{n+1} - \overline{\boldsymbol{\sigma}}_n - \mathbb{D}_0 : (\Delta\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{r}_{n+1}\Delta\phi_{n+1}) = \mathbf{0} \\ \kappa_{n+1}^p - \kappa_n^p - \Delta\phi_{n+1}h_{n+1}^p = 0 \\ \Delta\phi_{n+1} \geq 0, f_{n+1} \leq 0, \Delta\phi_{n+1}f_{n+1} = 0 \end{cases} \tag{41}$$

In the plastic calculation, the variables at step $n+1$, including $\overline{\boldsymbol{\sigma}}_{n+1}$, κ_{n+1}^p, and $\Delta\phi_{n+1}$, are obtained by solving Eq. (41), where the known conditions are the values of these variables at step n, as well as the strain increment $\Delta\boldsymbol{\varepsilon}_{n+1}$ of the current loading step. Following the return mapping stress update algorithm (Yu et al., 2024), the solution procedure of Eq. (41) is divided into two steps: elastic prediction and plastic correction, as shown in Fig. 4(a). First, the elastic trial stress is calculated through Hooke's law: $\overline{\boldsymbol{\sigma}}_{n+1}^{trial} = \overline{\boldsymbol{\sigma}}_n + \mathbb{D}_0 : \Delta\boldsymbol{\varepsilon}_{n+1}$. If $\overline{\boldsymbol{\sigma}}_{n+1}^{trial}$ lies inside the current yield surface, i.e., $f\left(\overline{\boldsymbol{\sigma}}_{n+1}^{trial}, \kappa_n^{\mathrm{p}}\right) \leq 0$, only elastic deformation occurs during the current loading step. $\overline{\boldsymbol{\sigma}}_{n+1}^{trial}$ is accepted as the true stress $\overline{\boldsymbol{\sigma}}_{n+1}$ at step $n+1$. The values of variables at step $n+1$ are obtained by

$$\begin{cases} \overline{\boldsymbol{\sigma}}_{n+1} = \overline{\boldsymbol{\sigma}}_n + \mathbb{D}_0 : \Delta\boldsymbol{\varepsilon}_{n+1} \\ \kappa_{n+1}^p = \kappa_n^p \\ \boldsymbol{\varepsilon}_{n+1}^p = \boldsymbol{\varepsilon}_n^p \\ \Delta\phi_{n+1} = 0 \end{cases} \tag{42}$$

If $\overline{\boldsymbol{\sigma}}_{n+1}^{trial}$ exceeds the current yield surface, i.e., $f\left(\overline{\boldsymbol{\sigma}}_{n+1}^{trial}, \kappa_n^{\mathrm{p}}\right) > 0$, it indicates that the material has undergone plastic flow. The following nonlinear equations need to be solved to conduct the plastic correction:

$$\{\boldsymbol{f}\}_{n+1} = \begin{Bmatrix} \overline{\boldsymbol{\sigma}}_{n+1} - \overline{\boldsymbol{\sigma}}_n - \mathbb{D}_0 : (\Delta\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{r}_{n+1}\Delta\phi_{n+1}) \\ \kappa_{n+1}^p - \kappa_n^p - \Delta\phi_{n+1}h_{n+1}^p \\ f(\overline{\boldsymbol{\sigma}}_{n+1}, \kappa_{n+1}^p) \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ 0 \\ 0 \end{Bmatrix} \tag{43}$$

Generally, the Newton-Raphson method is commonly employed to solve nonlinear systems of equations during plastic correction due to its quadratic convergence rate. However, the Newton-Raphson method may encounter convergence difficulties when the load increment is too large or the nonlinearity of the constitutive model is too strong. Herein, the line search method is employed to replace the Newton-Raphson method for solving Eq. (43). The fundamental concept of this method is to enhance the solution's convergence by making the search step size as optimal as possible. The iterative formula of the line search method is:

$$\{\boldsymbol{x}\}_{n+1}^{k+1} = \{\boldsymbol{x}\}_{n+1}^{k} + \alpha^k\{\boldsymbol{d}\}_{n+1}^{k} \tag{44}$$

where $\{\boldsymbol{x}\}_{n+1} = \left\{\overline{\boldsymbol{\sigma}}_{n+1} \quad \kappa_{n+1}^p \quad \Delta\phi_{n+1}\right\}^T$. $\{\boldsymbol{d}\} = -[\boldsymbol{J}]^{-1}\{\boldsymbol{f}(\boldsymbol{x})\}$ is the search direction. $[\boldsymbol{J}]$ is the Jacobian matrix and $[\boldsymbol{J}] = [\partial\boldsymbol{f}/\partial\boldsymbol{x}]$. The determination of $[\boldsymbol{J}]$ requires derivatives of the nonlinear stress integral equation system with respect to the independent variables. For the Mohr-Coulomb model, details of these derivative terms can be found in the literature (Zhou et al., 2023). α is the search step size, which is obtained through the following iterative formula (Zhou et al., 2022):

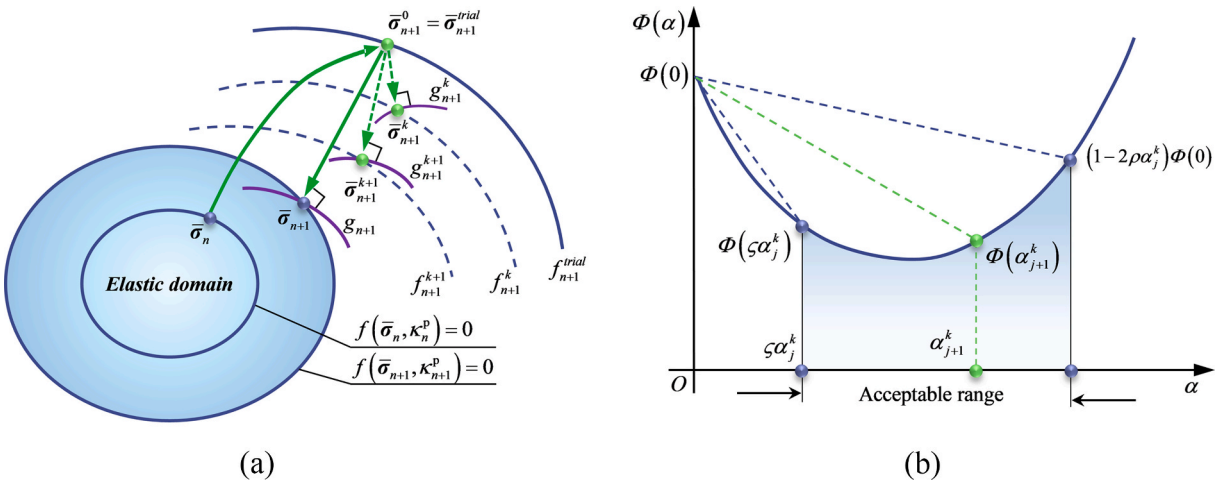


Fig. 4. Line search based-return mapping algorithm: (a) elastic prediction and plastic correction; (b) line search method.

$$\begin{cases} \text{Accept } \alpha_j^k \text{ and exit} & \text{IF } \Phi\left(\alpha_j^k\right) < \left(1-2\rho\alpha_j^k\right)\Phi(0) \\ \alpha_{j+1}^k = \max\left[\varsigma\alpha_j^k, \dfrac{\Phi(0)}{\Phi(0)+\Phi\left(\alpha_j^k\right)}\right] & \text{ELSE} \end{cases} \tag{45}$$

where ρ and ς are both algorithm parameters. Recommended values from the literature (Scherzinger, 2017) are 10^{-4} and 0.1, respectively. Φ is the merit function corresponding to the nonlinear system of equations Eq. (43).

$$\Phi\left(\alpha_j^k\right) = \frac{1}{2}\left\{\boldsymbol{f}\left(\boldsymbol{x}_n + \alpha_j^k\{\boldsymbol{d}\}_{n+1}^k\right)\right\}^T\left\{\boldsymbol{f}\left(\boldsymbol{x}_n + \alpha_j^k\{\boldsymbol{d}\}_{n+1}^k\right)\right\} \tag{46}$$

Eq. (45) employs an inexact line search technique that does not require the search step size to be the optimal value but rather a relatively optimal range for the current search direction, as shown in Fig. 4 (b). The standard Newton-Raphson method is a special case of the line search method when $\alpha = 1$.

3.3. Nonlocal damage calculation

In the damage calculation, the nonlocal damage internal variable defined by Eq. (1) is computed using the discrete form:

$$\widehat{\kappa}_{n+1}^{\mathrm{d}} = (1-m)\kappa_{n+1,i}^{\mathrm{d}} + m\frac{\sum_{j=1,j\neq i}^{NIP} v_j w_j \kappa_{n,j}^{\mathrm{d}} + v_i w_i \kappa_{n+1,i}^{\mathrm{d}}}{\sum_{j=1}^{NIP} v_j w_j} \tag{47}$$

where NIP represents the total number of neighbourhood integration points. v_j and w_j represent the element volume and weight value, respectively. The value of m needs to be greater than 1 to ensure the effectiveness of regularization, and should not be too large to cause potential numerical instability. The default value of $m = 1.1$ is used herein.

$\widehat{\kappa}_{n+1}^{\mathrm{d}}$ obtained by Eq. (47) will be further used to calculate the damage variable $\omega\left(\widehat{\kappa}_{n+1}^{\mathrm{d}}\right)$. The nominal stress at $n+1$ step is obtained by putting $\omega\left(\widehat{\kappa}_{n+1}^{\mathrm{d}}\right)$ into Eq. (27):

$$\boldsymbol{\sigma}_{n+1} = \left(1-\omega\left(\widehat{\kappa}_{n+1}^{\mathrm{d}}\right)\right)\bar{\boldsymbol{\sigma}}_{n+1} \tag{48}$$

Computing the nonlocal variable requires information about neighbourhood integration points. Generally, the traversal algorithm is widely used to search neighbourhood integration points due to its simple structure. However, the traversal algorithm is not suitable for large-scale numerical examples since its time complexity is $O(N)$. To improve the computational efficiency, we employ an Octree-nonlocal computation method, which consists of two acceleration approaches: (*i*) a local-nonlocal hybrid computing strategy. Nonlocal computation is initiated only during the softening stage, while those during the hardening stage adheres to the local computation. The threshold condition for nonlocal computation is set to $\varepsilon_{\mathrm{d}}^{\mathrm{p}} \geq 10^{-4}$. A similar acceleration strategy can be found in the literature (Mánica et al., 2018). (*ii*) The Octree algorithm. When performing radius neighbourhood search operations, its time complexity is $O(\log N + M)$ (Bédorf et al., 2012; Ram and Sinha, 2019), meaning that as the data size increases, the algorithm's running time grows more slowly. Here, N and M denote the total number of data points and the number of data points within the query radius, respectively. It is worth noting that in worst-case scenarios involving large search radius or unbalanced trees, octree may degrade to $O(N)$, while k-d tree maintains $O(N^{1-1/d} + M)$ (de Berg et al., 2008; Samoladas et al., 2022) in the d-dimensional space. Theoretically, the k-d tree algorithm may hold an advantage due to its lower time complexity in the worst case. However, for the segmentation and search of 3D data, the octree algorithm is simpler to implement and offers greater performance potential in practice. The research by Behley et al. (2015) demonstrates that by pruning irrelevant subtrees, the optimized octree algorithm achieves speedups of 1.2–2.7 × compared to the best k-d tree implementations across different datasets. Consequently, the octree algorithm was adopted in this work to handle neighbourhood search tasks in nonlocal computations.

The Octree, also known as the quadtree in 2D conditions. The octree algorithm's search procedure comprises two stages. *Step 1*: Constructing a tree data structure. First, construct a root node space (a cube for 3D cases or a square for 2D cases) that covers the entire computational domain. Then, recursively divide each node space into 4 equal parts for 2D cases (or 8 equal parts for 3D cases). The termination condition for division is defined arbitrarily. For example, in Fig. 5, the termination condition is that each node space accommodates no more than five integration points. Node spaces that are not further divided are called leaf node spaces, which store all the information of the integration points.

Step 2: Searching for the neighbourhood integration points. As shown in Fig. 6, the search initiates from the root node space to identify the leaf node spaces that intersect with the vicinity of the point of interest, known as the near-leaf node spaces. If a node space has no overlap with the neighbourhood of the current integration point, such as node spaces ① and ② in Fig. 6, then this node space and its child node spaces will be directly excluded. In practical calculations, to facilitate the determination of whether the neighbourhood overlaps with the node space, the circular neighbourhood can be replaced with a square that is adjacent to it. Finally, the neighbourhood integration points are acquired by

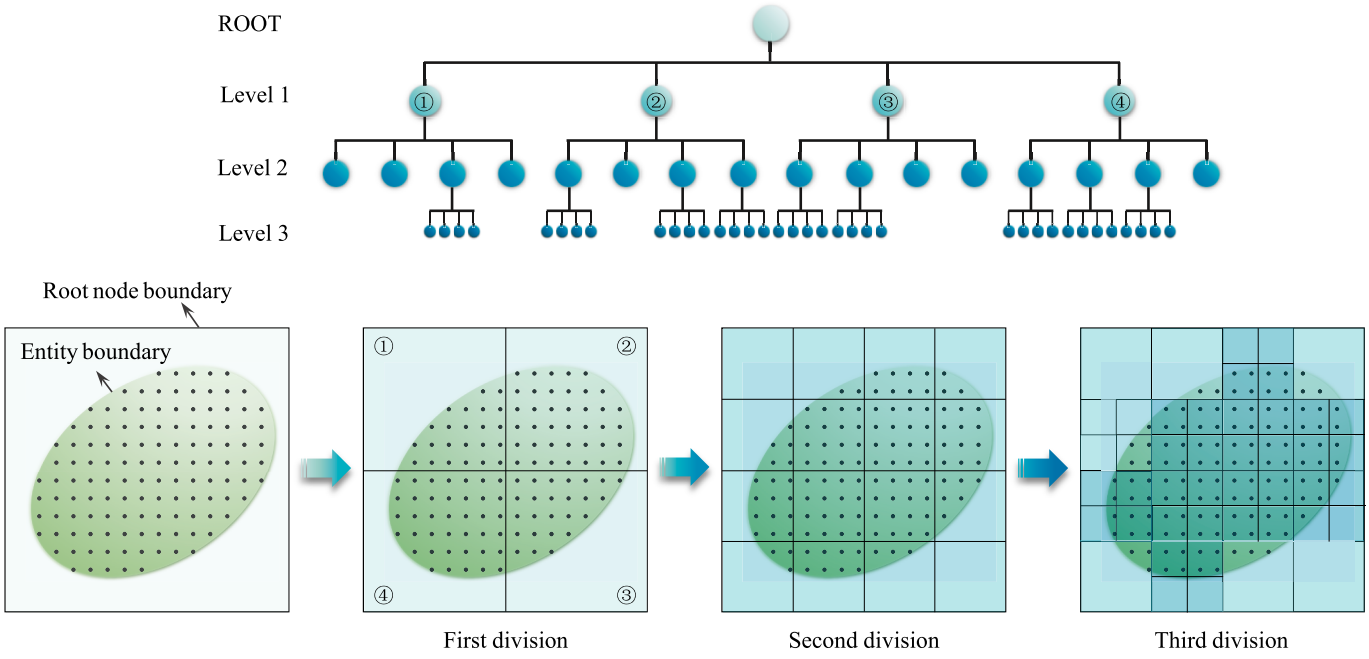


Fig. 5. Construction of the tree data structure.

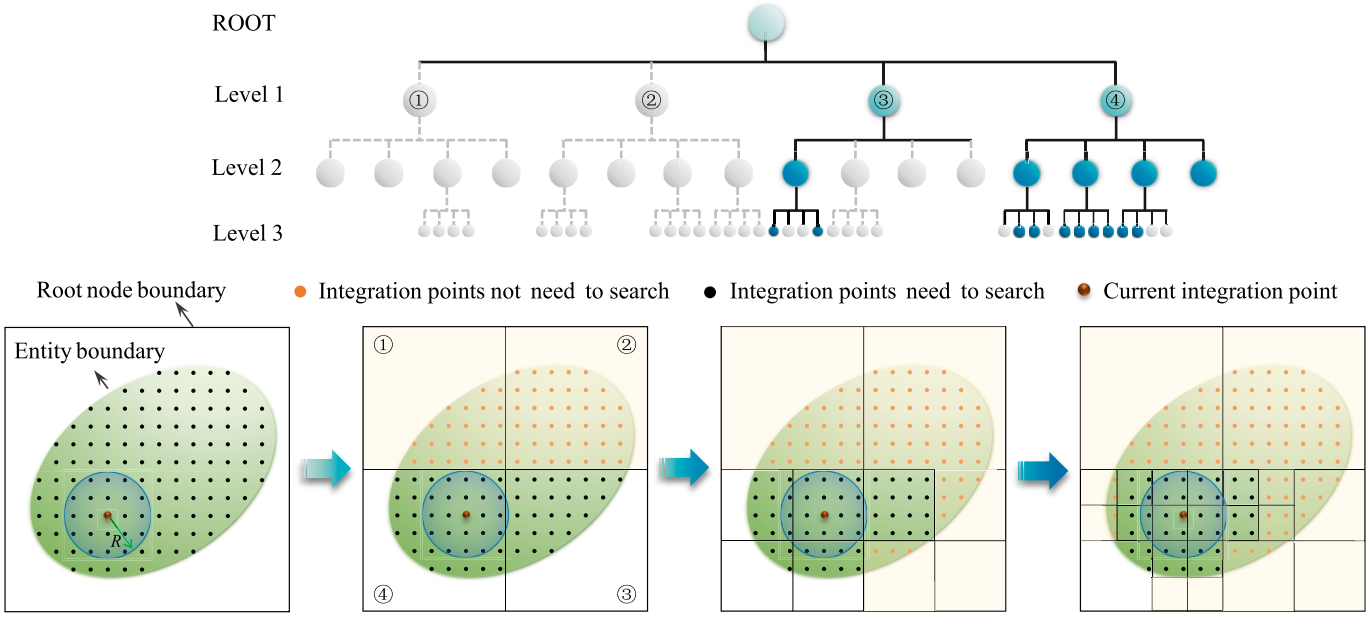


Fig. 6. Neighbourhood search process based on the tree data structure.

evaluating the distances between the current integration point and integration points in near-leaf node spaces. It is worth noting that in the octree search algorithm, we first determine if the node-space overlaps with the neighbourhood of the interest point. If so, further judgment is performed to confirm if its child node spaces and corresponding integration points indeed fall into the neighbourhood of interest point. Although node space ③ does not contain nodes within the neighbourhood of the current integration point directly, there is a minor overlap between the two. Therefore, a search task needs to be executed for node space ③.

3.4. Numerical implementation

The nonlocal Mohr-Coulomb plasticity damage model and octree-nonlocal computation method have been implemented into the ABAQUS software via the UMAT subroutine. Unlike the numerical implementation of local models, nonlocal models require accessing information from neighbourhood integration points to calculate the nonlocal variable, as shown in Eq. (47). To enable this nonlocal calculation, we have customized two FORTRAN modules, named GlobalVar and OctreeAlg, as shown in Fig. 7(a).

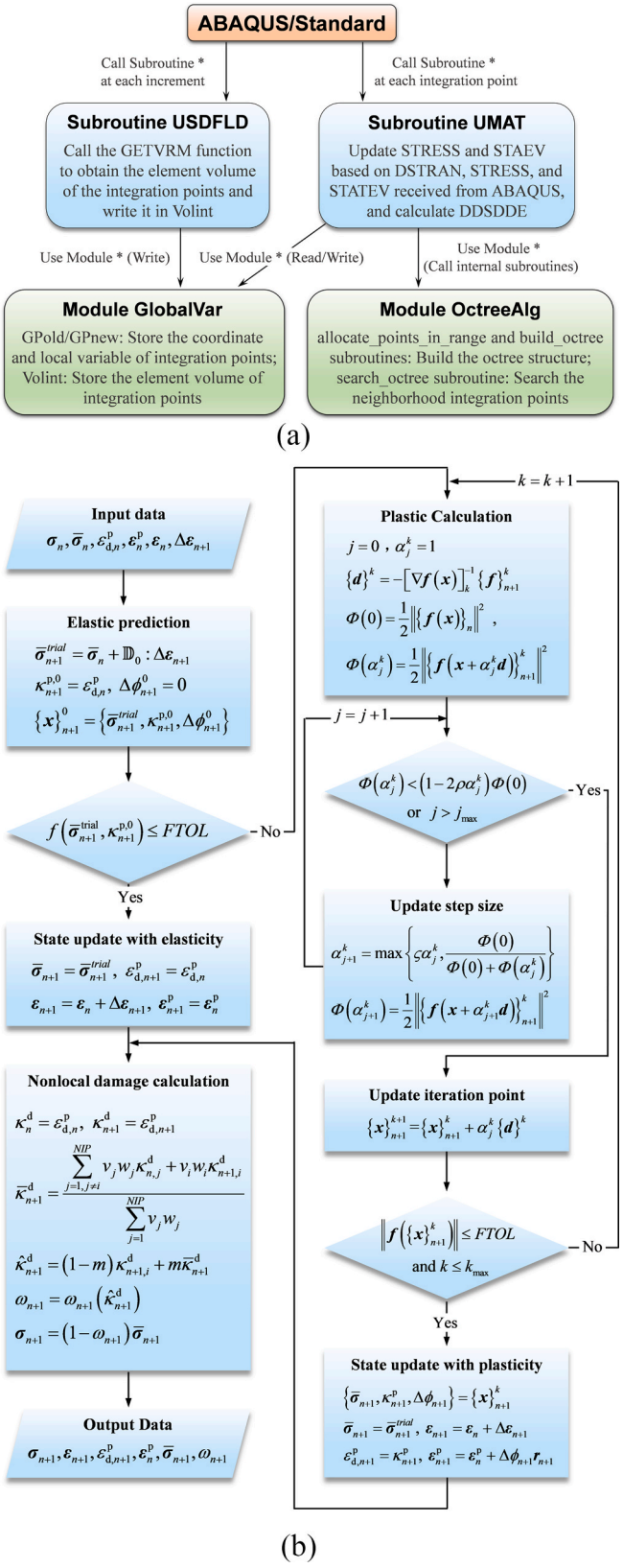


Fig. 7. Numerical implementation of the nonlocal model: (a) general flow; (b) stress update process.

The octree algorithm used to search neighbourhood integration points is encapsulated in the OctreeAlg module. It encapsulates the necessary data types and core functionalities for the algorithm. Inside the OctreeAlg module, we utilize FORTRAN's Type statement to customize two data types: "point" and "node". The "point" type stores essential information for a single integration point, such as coordinates and local variable. The "node" type describes the node space of the octree, containing the spatial domain it represents (dimensions, position), pointers to its child node spaces and parent node spaces (establishing the hierarchical structure), and a list of integration points residing within its spatial bounds. In addition, the module defines three internal subroutines using the Contains keyword. Subroutine allocate_points_in_range(…) and build_octree(…): These two subroutines are responsible for constructing the octree data structure. It recursively partitions the computational domain space into smaller cubes for the 3D cases or squares for the 2D cases and distributes the integration points into their final leaf node spaces, continuing until predefined termination criteria are met. It is worth emphasizing that octree reconstruction occurs at each increment's start, using integration point coordinates from the current configuration if NLGEOM=ON (otherwise, initial coordinates). Thus, this method is not only applicable to small strains but also to finite strains. Subroutine search_octree(…): This subroutine performs an efficient search for the neighbourhood integration points. Starting from the root node space, it leverages the spatial relationship between the target neighbourhood and the node spaces to rapidly exclude vast regions of irrelevant node spaces. It ultimately returns only the leaf nodes that intersect with the target neighbourhood. The final neighbourhood points are then determined by examining the integration points stored within these few relevant leaf nodes, dramatically reducing the search costs compared to a global traversal algorithm. Within the UMAT subroutine, this encapsulated module is accessed via the Use OctreeAlg statement. Specifically, the allocate_points_in_range and build_octree subroutines are called, typically at the start of an increment or as needed, to build or update the octree structure. The search_octree subroutine is called to perform the efficient spatial search. Both the construction logic and the search queries are fully encapsulated within the OctreeAlg module and invoked directly by UMAT.

In the GlobalVar module, three global arrays (with the SAVE attribute by default) are declared: GPold(NEL, NIP, 4) stores the information (coordinates and local variables) for all integration points in the whole computation domain at step n; GPnew(NEL, NIP, 4) is used to record the information for these points at step $n+1$; and Volint(NEL, NIP, 1) stores the element volume associated with each integration point, obtained via the USDFLD subroutine. Here, NEL is the total number of elements, and NIP is the number of integration points per element. At the start of each new increment step, the operation GPold = GPnew is performed once. This updates the reference state to the converged state from the end of the previous increment. During subsequent UMAT calls within the increment, the GPnew array is progressively updated for the currently processed integration point with its value of the local variable at step $n+1$. When computing the nonlocal variable by Eq. (47), only the local variable of the current integration point utilizes the value at step $n+1$. In contrast, the local variables of all other neighbourhood integration points utilize values at step n, accessed via the GPold array. It can be found that there is a temporal discrepancy for the semi-implicit integration scheme employed in Eq. (47). This design is a deliberate choice made to balance computational efficiency with accuracy – a common consideration in implementing nonlocal integral models. Crucially, it avoids the tedious interaction between different integration points caused by iterative updates of local variables within one time step. The resulting error is manageable and acceptable given that its magnitude

scales with the size of the time step, especially when the small time increments are used. This approach has been successfully applied in implementations of nonlocal models, as seen in references (Mánica et al., 2018; Rolshoven, 2003). Fig. 7(a) illustrates the main functions, data flow, and call sequence of the subroutines and modules described above. Fig. 7(b) details the stress update process of the nonlocal Mohr-Coulomb plasticity damage model.

FORTRAN provides two ways to define global variables, namely COMMON blocks and Modules. Modules were chosen over COMMON blocks for defining these global arrays primarily due to their support for allocatable arrays and encapsulation, which significantly enhance code maintainability. The SAVE attribute of module variables ensures their persistence between UMAT and other subroutine calls. The UMAT code for any given integration point accesses the global data via Use GlobalVar.

It should be emphasized that the softening curve is a post-bifurcating branch. Mathematically, bifurcation signifies the emergence of multiple solutions satisfying the equilibrium equations and boundary conditions. In practice, material imperfections or other inhomogeneities lead to the selection of a single solution from these possibilities (Varas et al., 2005). Consequently, from the perspective of bifurcation theory, the softening curve depends on initial imperfections and perturbations (de Borst, 1987). This work does not challenge these physical foundations nor claim "constitutive reality" for softening behaviour. Given the complexity and path-dependent microstructural evolution of geomaterials, developing universal first-principles constitutive equations currently remains impractical. Within this context, phenomenological softening models embed the softening branch into governing equations as a pragmatic and relatively simple approach. Widely adopted models, e.g., Sanisand model (Taiebat and Dafalias, 2008) and Concrete damaged plasticity model (Lee and Fenves, 1998), validate the practical utility of this methodology. It provides tractable computational models for describing macroscopic behaviours of geomaterials, serving engineering analyses such as foundation bearing capacity, slope failure analysis, and tunnel stability assessment.

The core contribution of this work lies in enhancing the numerical reliability and computational efficiency of these practical phenomenological models. While conventional nonlocal methods mitigate mesh dependency, the regularization parameter (i.e., the characteristic length l_c) still significantly biases predictions of structural softening responses. The proposed scaled nonlocal method directly addresses this issue. During our investigation of the 1D nonlocal response in a rod under tension, we observed that the distribution of plastic strain satisfies a Fredholm integral equation of the second kind, which furnishes a solid analytical foundation for exploring how the characteristic length influences the nonlocal response. From the derived 1D analytical solution, we found that variations in characteristic length induce proportional scaling of load–displacement response along the plastic displacement axis. Exploiting this insight, we proposed a scaling expression (i.e., Eq. (20)) to eliminate the influence of the characteristic length on the predicted structural softening response, thereby ensuring uniqueness and objectivity regardless of mesh size or characteristic length. Subsequently, this scaled nonlocal method is applied to a Mohr–Coulomb plasticity damage model. The model is further implemented numerically via an efficient nonlocal computation method, which integrates an octree-based neighbourhood search technique and a local–nonlocal hybrid computation strategy to overcome the computational bottlenecks inherent in nonlocal models. Although the scaled nonlocal method is grounded on analytical derivations under 1D conditions, its effectiveness in 2D and 3D settings will be thoroughly validated via numerical examples presented in Section 4. It is important to emphasize that the purpose of this study is not to define a physical meaning for the characteristic length, but rather to mathematically eliminate its interference with the structural softening response. This approach can significantly enhance the practicality of nonlocal models for failure analysis of engineering structures, freeing researchers from concerns related to mesh resolution and characteristic length selection.

4. Numerical examples

This section systematically evaluates the performance of the scaled nonlocal method through three benchmark boundary value problems: (1) 3D bar tension, (2) plane strain compression, and (3) slope instability analysis. These examples consider different element sizes, characteristic lengths, and forms of softening functions. Particular emphasis is placed on evaluating the efficiency of the Octree-nonlocal computation method in the last two examples. All numerical experiments were conducted on the same workstation with an AMD Ryzen 9 7950X 16-core Processor (4.50 GHz) and 64 GB of RAM, with model parameters detailed in Table 1.

4.1. Tension problem of a 3D bar

The first benchmark case examines a 3D bar (length 150 mm, cross-section 1 mm × 1 mm) under uniaxial tension, where the left end is horizontally constrained and the right end undergoes 1 mm prescribed displacement, as shown in Fig. 8. To induce strain localization, a cubic material imperfection zone with 99.5 % reduced cohesive strength is strategically embedded. The finite element model employs uniformly distributed 8-node linear brick elements (C3D8) with a mesh size $l_{FE} = 1$ mm. In this example, three softening laws: linear softening, exponential softening, and rational fraction softening, as well as three characteristic lengths $l_c = 5$ mm, 10 mm, and 15 mm, are considered. The scaling factor is set to $\eta = l_c/(10\ \text{mm})$, which ensures that the scaled nonlocal model's load–displacement responses across different characteristic lengths will match the original nonlocal model's solution at $l_c = 10$ mm.

Fig. 9 compares the load–displacement responses of both nonlocal models under linear softening. The original model exhibits strong characteristic length dependence. As l_c increases, the load–displacement curves' softening rate continuously drops. In contrast, the load–displacement response from the scaled nonlocal model is insensitive to l_c. All curves ($l_c = 5/10/15$ mm) collapse onto the original model's solution at $l_c = 10$ mm.

Fig. 10(a) and (b) compare the damage distribution nephogram between both nonlocal models, with the strain localization boundary (defined by $\omega = 10^{-4}$) marked by dark blue dashed lines. It is observed that strain localization phenomena are distributed near the weak area. The localization zone width predicted by both nonlocal models is essentially the same under the same l_c and increases proportionally with l_c. This result is consistent with the conclusion of Eq. (A39): the localization zone width is controlled by the characteristic length. Additionally, within the localization zone, the damage degree predicted by the original nonlocal model is higher at $l_c = 5$ mm and lower at $l_c = 15$ mm than that calculated by the scaled nonlocal model. The difference in the damage distribution also suggests that the load–displacement response with the original nonlocal model softens faster at $l_c = 5$ mm and slower at $l_c = 15$ mm than that with the scaled nonlocal model.

Fig. 11 and Fig. 12 systematically demonstrate the scaled nonlocal method's effectiveness under exponential and rational fraction softening laws. Similar to the results presented in Fig. 9, as l_c increases, the

Table 1
Model parameters used in the boundary problems.

Boundary value problems	E/MPa	ν	c_0/kPa	φ/°	ψ/°	A_s
Tension problem of a 3D bar	30,000	0.3	20000	20	0.1	20, 20, 2000
Plane strain compression of a rectangle plate	60,000	0.3	10000	20	0.1	4000
Stability analysis of strip footing on the slope	20	0.3	8	10	3	8

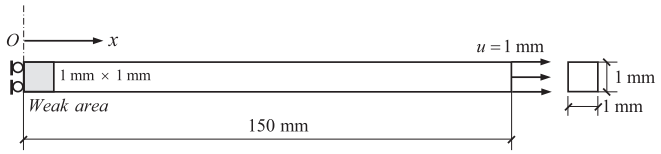


Fig. 8. Tension problem of a 3D bar.

softening response obtained by the original nonlocal model gradually decreases, showing severe characteristic length sensitivity. Conversely, under different characteristic length conditions, all curves obtained by the scaled nonlocal model completely overlap. Presented in the simulation results of Fig. 9, Fig. 11, and Fig. 12 also indicate that the proposed scaled nonlocal method is valid for any form of softening law. The influence of the characteristic length on the load–displacement response is effectively counteracted by inverse scaling the nonlocal variable $\hat{\kappa}^{\mathrm{d}}$, preserving solution objectivity without empirical regularization.

4.2. Plane strain compression test

The compression problem of a rectangular plate under the plane strain condition is a benchmark example for evaluating the effectiveness of the nonlocal model. As shown in Fig. 13, the plate is 60 mm long and

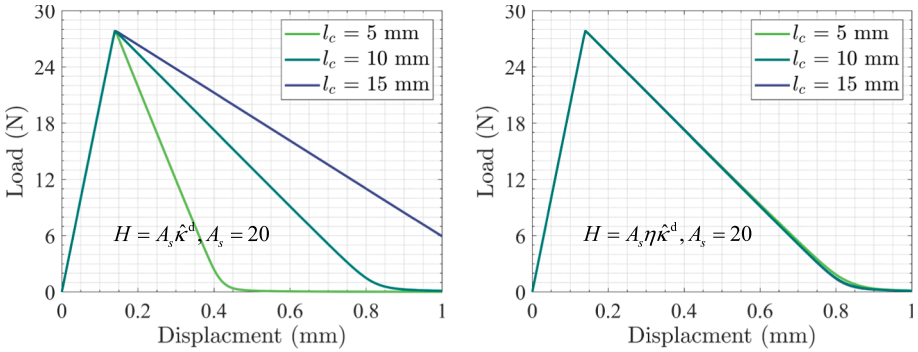


Fig. 9. Load displacement curves of the bar with linear softening law: (a) original nonlocal model; (b) scaled nonlocal model.

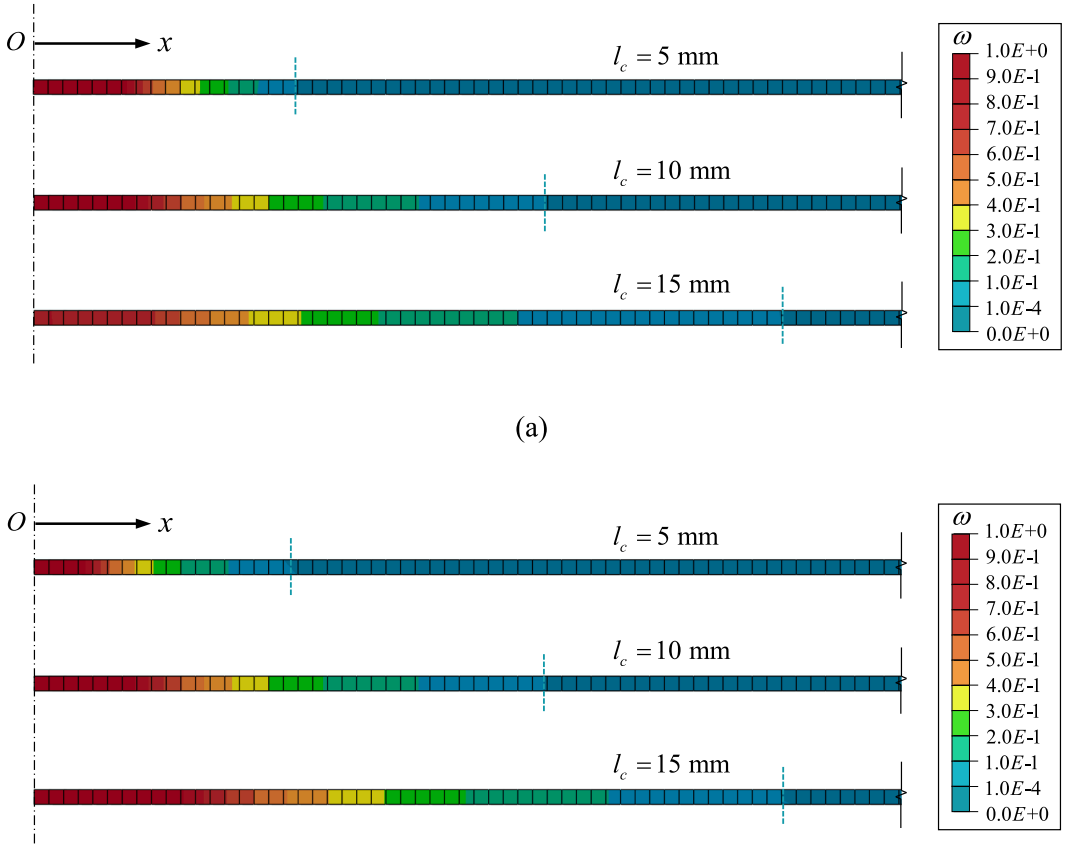


Fig. 10. Nephogram of damage variable of bar: (a) original nonlocal model; (b) scaled nonlocal model.

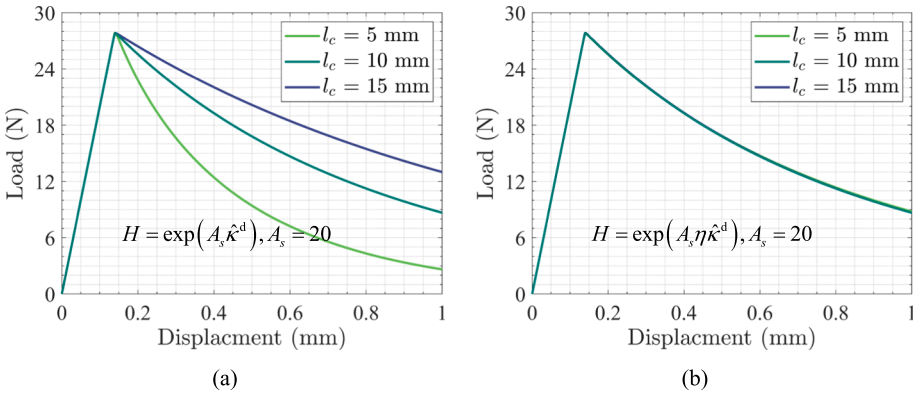


Fig. 11. Load displacement curves of the bar with exponential softening law: (a) original nonlocal model; (b) scaled nonlocal model.

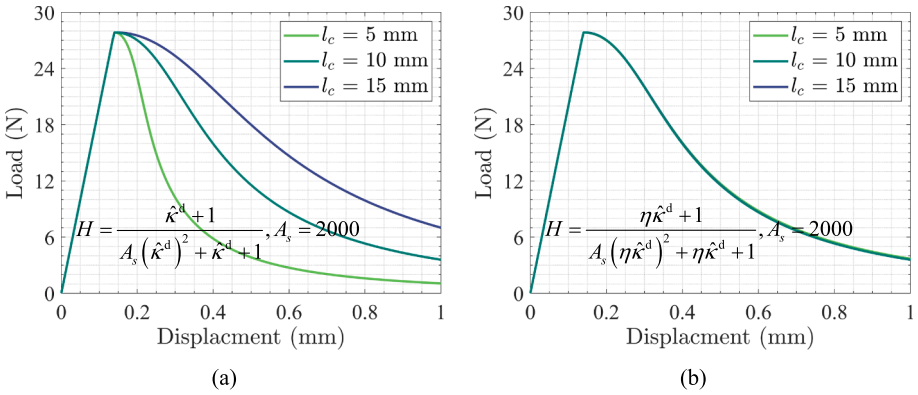


Fig. 12. Load-displacement curves of the bar with rational fraction softening law: (a) original nonlocal model; (b) scaled nonlocal model.

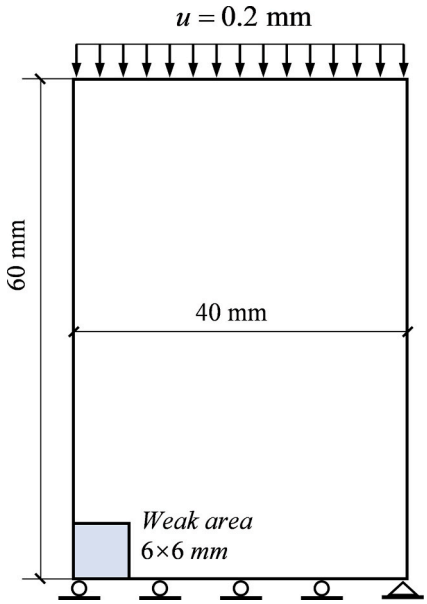


Fig. 13. Geometry and boundary conditions of 2D plate.

40 mm wide. The bottom end of the plate is constrained vertically. The bottom-right corner is constrained horizontally to prevent overall movement of the plate. The top of the plate is subjected to a 0.2 mm vertical displacement load. A 6 mm × 6 mm weak area at the bottom left of the plate initiates strain localization. For simplification, only the rational fraction softening law is used in this example.

Fig. 14 and Fig. 15 show simulation results from the local and nonlocal models. Spatial discretization uses 4-node plane strain elements (CPE4) across four mesh densities (600–6700 elements). The characteristic length of the nonlocal model is set to $l_c = 2.5$ mm. Inevitably, with mesh refinement, local simulations exhibit a continuous reduction in shear band width and an increase in softening response, exhibiting pathological mesh dependence. Conversely, the nonlocal formulation maintains a very small variation in both shear band width and load–displacement curves across all discretization levels, demonstrating consistent regularization capability. In what follows, let's discuss the shear band orientation under different mesh densities. Based on the force equilibrium analysis, Coulomb theory defines the upper bound of the shear band angle relative to the minor principal stress direction ($\theta_C = 45^\circ + \varphi/2$) suitable for fine-grained materials. Roscoe theory establishes the lower bound ($\theta_R = 45^\circ + \psi/2$) suitable for coarse-grained materials through the plastic flow rule. Vermeer (1990) further demonstrated that under plane strain conditions, actual shear band angles lie between these theoretical bounds, modulated by particle size distribution and boundary constraints. Fig. 14 results show that the shear band orientation predicted by the local model changes from 45° to 47° for different mesh densities, while the one predicted by the nonlocal model remains around 45°. This simulation result closely approaches the Roscoe bound and aligns with DEM (Discrete Element Method) observations by Hazeghian and Soroush (2022) and predictions from the nonlocal plasticity model presented by Mánica et al. (2017). However, compared to the local model, the shear band orientation predicted by the nonlocal model is less sensitive to method size.

The characteristic length sensitivity analysis employs four values: l_c = 1.0 mm, 1.5 mm, 2.0 mm, and 2.5 mm. The original nonlocal model's result with $l_c = 1.0$ mm is considered as the baseline by setting the scaling factor $\eta = l_c/(1.0\ \mathrm{mm})$. A mesh density of 6700 elements is

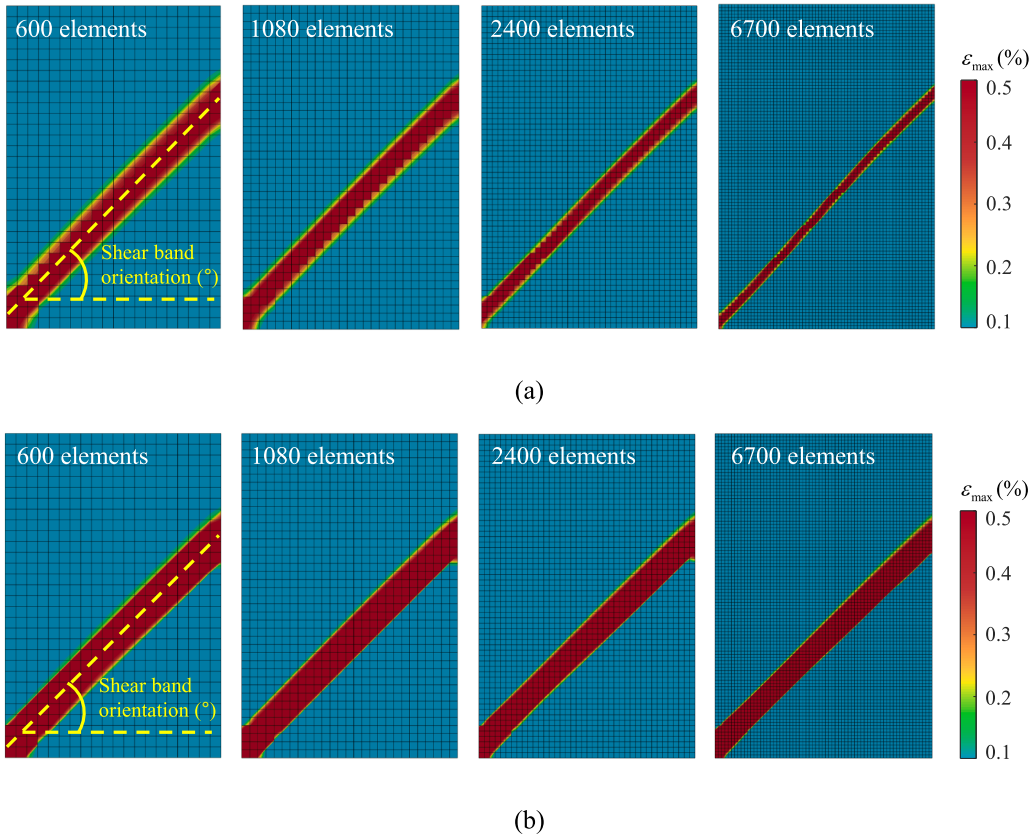


Fig. 14. Shear bands under different element sizes: (a) local model; (b) scaled nonlocal model.

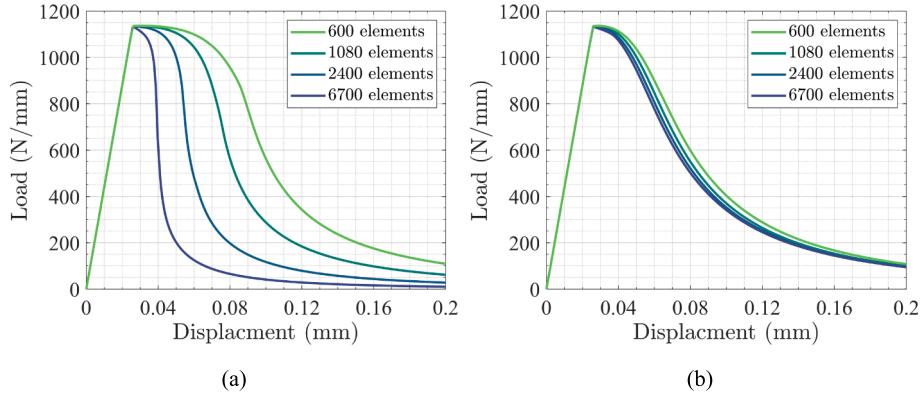


Fig. 15. Load displacement curves under different element sizes: (a) local model; (b) scaled nonlocal model.

selected to guarantee that the element size is smaller than the minimum characteristic length. Fig. 16 shows the shear bands characterized by damage localization. As l_c increases, the widths of the shear bands predicted by both nonlocal models gradually increase. For the original nonlocal model, the degree of damage localization inside the shear band gradually decreases with increasing characteristic length, while it remains essentially unchanged for the scaled nonlocal model.

Fig. 17 illustrates the load–displacement curves from both nonlocal models. The simulation results of the original nonlocal model significantly depend on l_c. As l_c increases, the softening response gradually decreases, consistent with the changing trend of damage degree inside shear bands illustrated in Fig. 16(a). When the scaled nonlocal model is used, the predicted load–displacement curves are insensitive to l_c and overlap with that of the original nonlocal model under the condition of $l_c = 1.0$ mm.

In what follows, the computational efficiency of the proposed octree-nonlocal method is compared against the traditional nonlocal method using the Mohr-Coulomb plasticity damage model presented in Section 3. CPU time consumption for both methods is recorded across numerical examples with varying numbers of elements. There are two key distinctions between the two nonlocal computation methods: (*i*) Neighbourhood search: The traditional method employs a traversal algorithm to search neighbourhood points, whereas the proposed method employs an octree algorithm. The octree algorithm partitions the computational domain into node spaces and evaluates their spatial relationships with the target neighbourhood, efficiently excluding irrelevant integration points during neighbourhood searches. (*ii*) Computation Strategy: The proposed method adopts a local-nonlocal hybrid computation strategy, activating nonlocal computations only within material damage softening zones. In contrast, the traditional method performs nonlocal

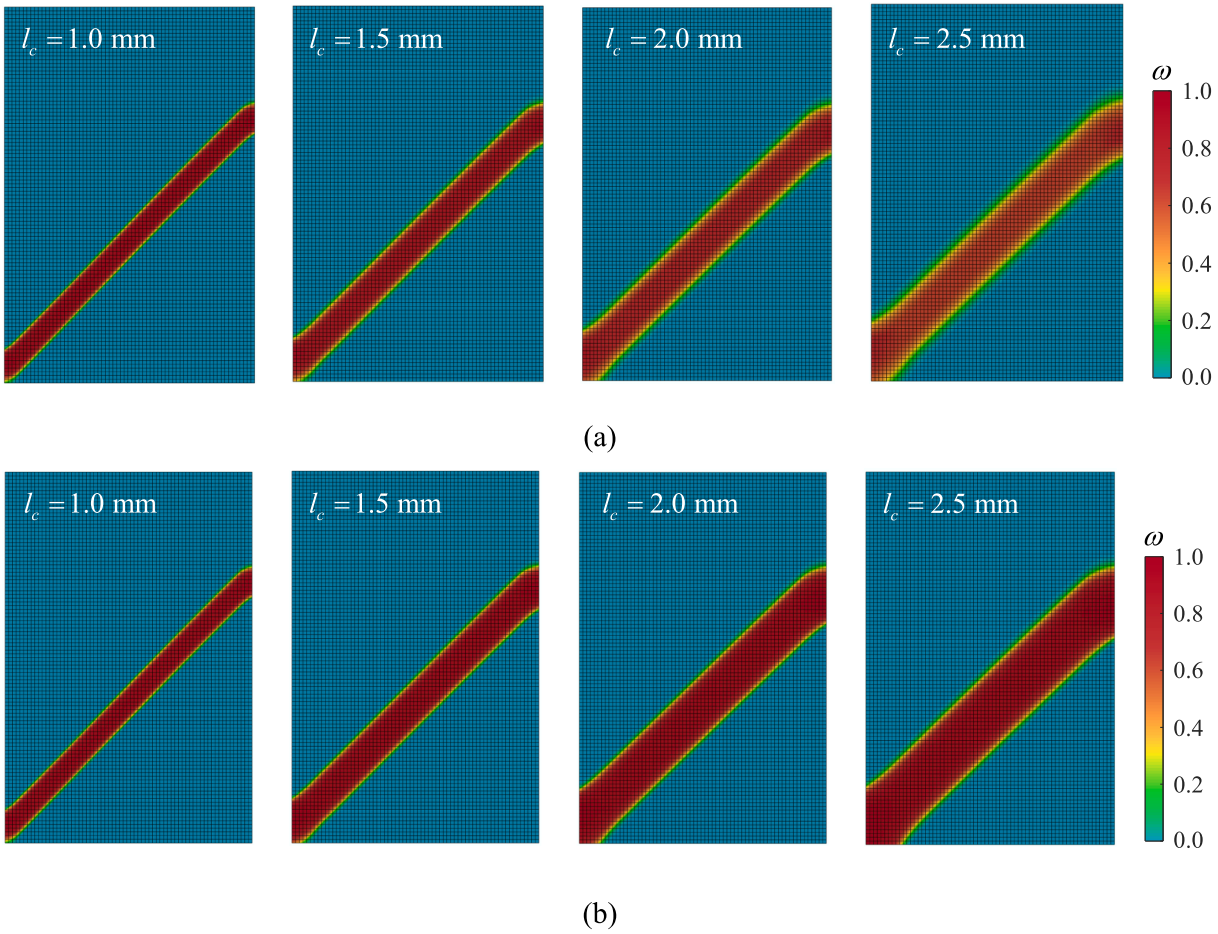


Fig. 16. Shear bands under different characterize lengths: (a) original nonlocal model; (b) scaled nonlocal model.

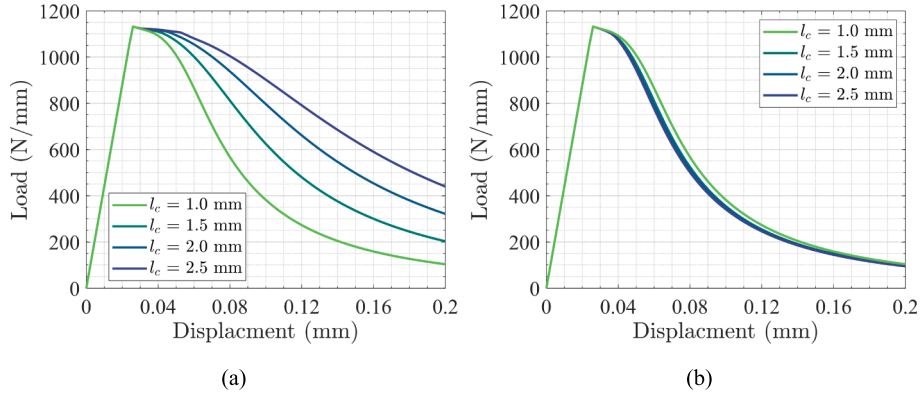


Fig. 17. Load displacement curves under characteristic lengths: (a) original nonlocal model; (b) scaled nonlocal model.

computations throughout the entire computational domain.

Fig. 18 shows the CPU times consumed by two methods as well as their ratio under different numbers of elements. The results show that the efficiency advantage of the proposed method is less significant for smaller element numbers due to the octree construction overhead. However, as the number of elements increases, its efficiency gain becomes markedly pronounced. The CPU time ratio between the two nonlocal methods increased from 1.41 to 4.67 times across meshes ranging from 600 to 6700 elements, indicating that the Octree-nonlocal computation method significantly improves nonlocal models' computational efficiency, especially for large-scale numerical examples.

4.3. Slope stability with strip footing

As the last example, we consider an instability problem of strip footing on a slope. As shown in Fig. 19, the slope is 12 m in length, 3.5 m in width, and has a slope ratio of 2.5. A 2 m wide rigid strip footing is positioned at the top of the slope. Although it is a plane strain problem, we still employ 3D finite element modelling to demonstrate the effectiveness of the proposed methods under 3D circumstances. The bottom of the slope is fully constrained, with horizontal displacements fixed on both sides. The initial stress in the slope is generated by the gravitational load, with a soil weight of $\gamma = 20\text{kN/m}^3$. During the loading process, the

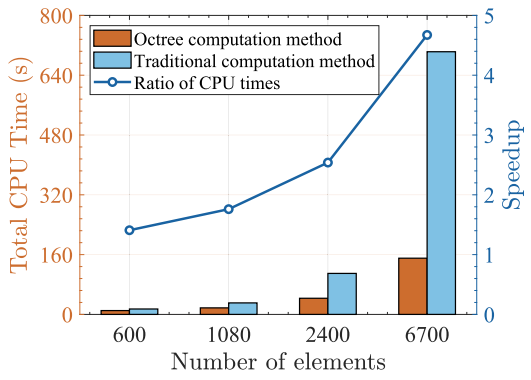


Fig. 18. Efficiency comparison of two computation methods in the plane strain compression test.

bottom of the rigid footing is subjected to a vertical displacement of $u = 0.06$ m, simulating the load applied by buildings on the slope.

In this example, only exponential softening laws are considered. Both the characteristic length and the element size are simultaneously varied, maintaining a constant ratio, i.e., $l_c/l_{FE} = 1.5$. The scaling factor is set to $\eta = l_c/(0.12\text{ m})$. Fig. 20 shows the shear bands obtained by both the local and scaled nonlocal models. Shear bands develop at the base of the rigid foundation on both sides, with those originating from the right side predominating and leading to the instability failure of the slope and footing. As l_{FE} decreases, the widths of the shear bands predicted by the two models decrease. However, the underlying causes differ: in the local model, this reduction is attributed to mesh dependence, whereas in the scaled nonlocal model, it results from the decrease of l_c. It should be emphasized that, in the current scaled nonlocal method, l_c acts merely as a numerical parameter influencing shear band width, rather than an intrinsic material property. On the other hand, experimental studies (Rechenmacher, 2006) and DEM simulations (Zhuang et al., 2014) have revealed that the shear bands of granular materials depend on density, particle shape/size, loading paths, and boundaries, as well as their widths and orientations developed and evolved during shearing (Amirrahmat et al., 2019; Finno et al., 1997), meaning that l_c may not a fixed parameter. For the scaled nonlocal method, when l_c is required to describe the width of shear bands observed in experiments, it may be considered to establish an intrinsic connection between l_c and these influencing factors and evolutionary mechanisms.

Fig. 21 shows the corresponding load–displacement curves. In the local simulations, the predicted peak load gradually decreases, and the post-peak softening curve becomes steeper as the mesh gets finer. Therefore, the local model cannot provide objective prediction results for failure analysis at the structural level. The load–displacement curves obtained from the scaled nonlocal model are insensitive to both element size and material characteristic length, demonstrating excellent consistency across different conditions.

Fig. 22(a) compares CPU time consumption between the proposed Octree-based nonlocal method and the traditional nonlocal method across varying numbers of elements. In this analysis, the total loading time and increment size were set to 1 s and 0.005 s, respectively, resulting in 200 load increments. As the number of elements increased from 2,583 to 36,057, the CPU time ratio of the traditional nonlocal computation method to the Octree-nonlocal computation method increased from 2.59 to 14.14. This demonstrates that the Octree-nonlocal computation method offers superior acceleration performance, even achieving an one order-of-magnitude improvement for large-scale simulations. A further comparison of computational efficiency between nonlocal and local models is provided in Fig. 22(b). Interestingly, for the case with 2583 elements, the CPU time for the nonlocal model exceeds that of the local model. However, as the element number increases, the CPU time ratio between local and nonlocal models increases from 0.86 to 1.40. This result challenges the conventional belief that local models typically offer significantly higher computational efficiency compared to nonlocal models. This counter-intuitive outcome is further clarified by the analysis of the global equilibrium iterations for both the local and nonlocal models, as shown in Fig. 23.

At the beginning of the loading process, the total number of global equilibrium iterations for both the local and nonlocal models is about the same. As deformation progresses, the number of global equilibrium iterations for local models rapidly increases, significantly exceeding those required for nonlocal models. Specifically, as the number of elements increases, the local model demands more iterations to achieve convergence of the global equilibrium equations. This is primarily due to the intensification of strain concentration within the shear band, which becomes more pronounced with mesh refinement. In contrast, the number of iterations for the nonlocal model remains relatively stable across different element counts. Therefore, although the local model has higher computational efficiency than the nonlocal model in each individual iteration step, the larger total number of global equilibrium iterations causes the local model to consume more CPU time compared to the proposed scaled nonlocal model. It is worth emphasizing that the nonlocal model in Fig. 23 adopts the Octree-nonlocal computation method presented in Section 3. The CPU time required by the traditional nonlocal computation method is still much higher than that of the local model.

5. Conclusions

This study presents a scaled nonlocal method that successfully decouples load–displacement predictions from both mesh density and characteristic length parameters. By deriving an analytical solution for a 1D bar problem, we rigorously demonstrate that characteristic length variations proportionally scale the load-plastic displacement curves, artificially altering the apparent softening rate of the structures. Leveraging this insight, we introduce a dimensionless scaling factor into

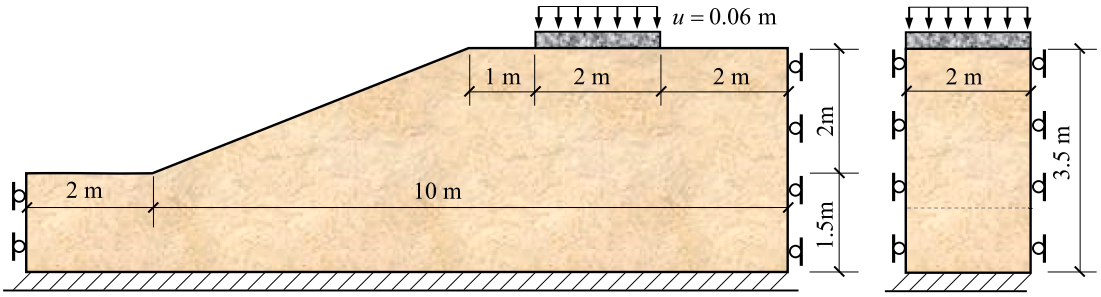


Fig. 19. Boundary condition and geometry profile of the slope.

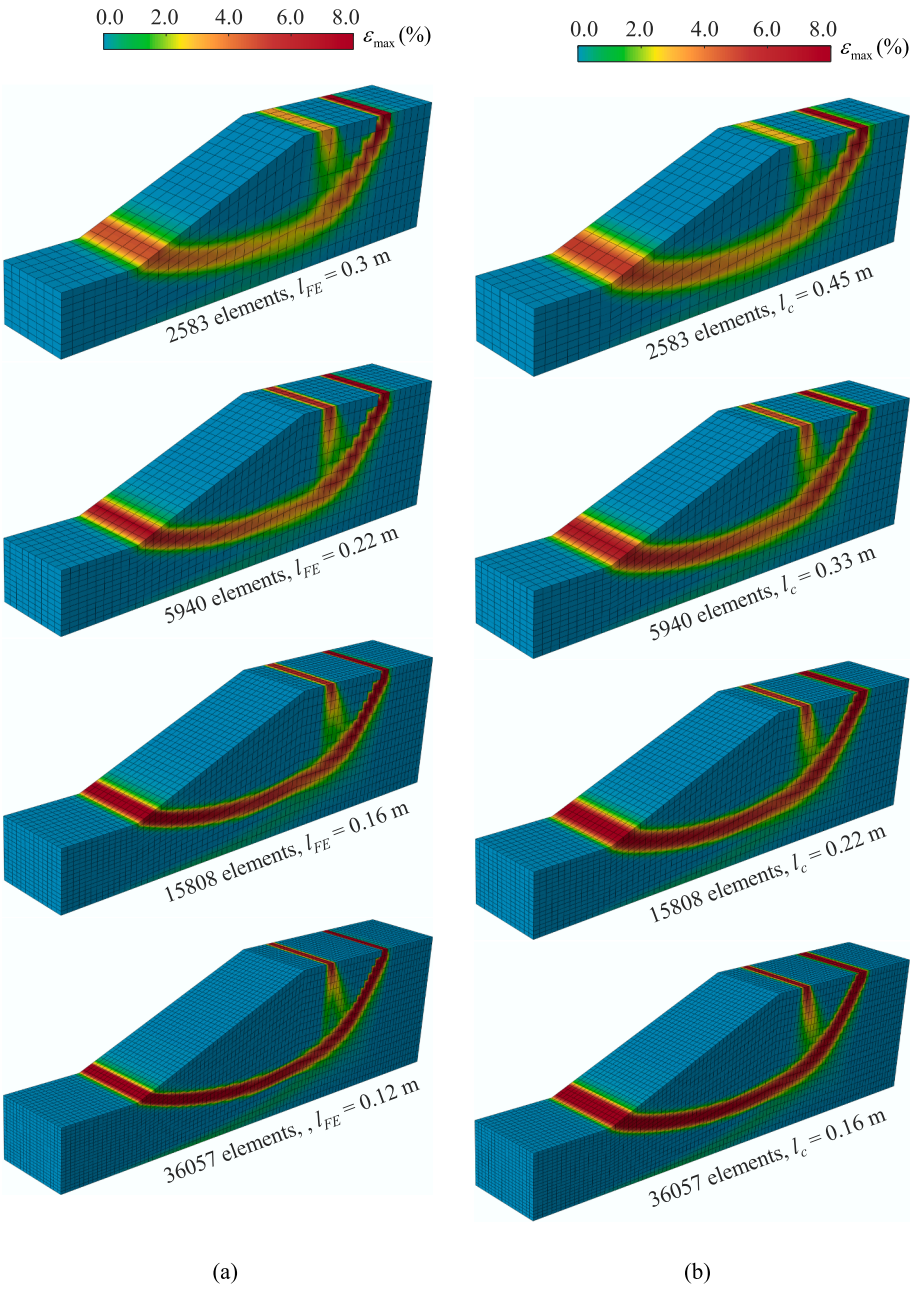


Fig. 20. Shear bands of slope under different element sizes: (a) local model; (b) scaled nonlocal model.

constitutive relations to neutralize characteristic length dependence and further integrate this theoretical advancement into ABAQUS UMAT implementation of a Mohr-Coulomb plasticity damage model.

Numerical validation across three boundary value problems, tested under diverse discretizations, characteristic lengths, and softening functions, confirms the robustness of the proposed method. Our results reveal that conventional nonlocal models exhibit characteristic length-dependent softening rates, whereas the scaled formulation achieves mesh- and length-insensitive responses regardless of softening law specifics. Critically, the characteristic length retains sole control over localization zone width, preserving its physical interpretability. To address computational bottlenecks, we develop an Octree-nonlocal computation method combining the local-nonlocal computation strategy (nonlocal computation activated only at softening integration points) with an octree-based neighbourhood search technique. This computation method achieves 2.59 to 14.14 times speedups over the traditional nonlocal method as element counts scale from 2,583 to 36,057, with efficiency gains surpassing even local models in large-scale simulations due to reduced global equilibrium iterations.

Derived from a 1D analytical solution, the scaled nonlocal method currently applies only to scalar-valued internal variables, with generalization to tensor-valued variables and rigorous proofs to higher dimensions requiring further investigation. However, its consistent accuracy in 2D/3D benchmarks highlights its potential for broader

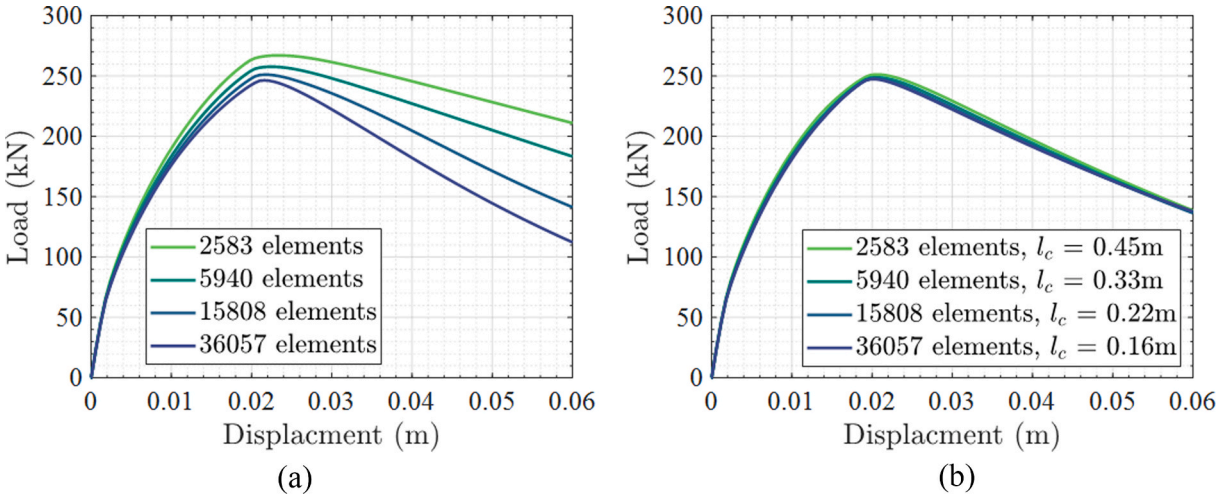


Fig. 21. Load displacement curve of strip footing: (a) local model; (b) scaled nonlocal model.

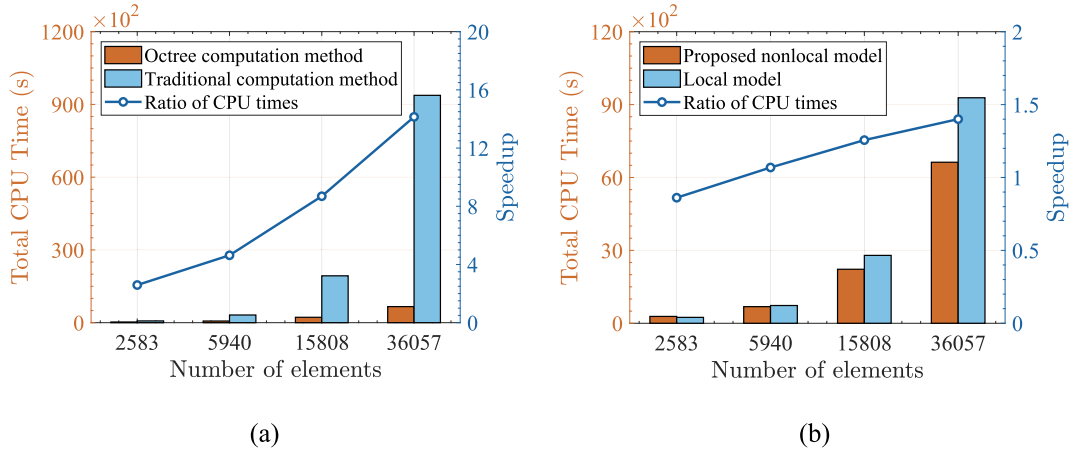


Fig. 22. Efficiency comparison of different computation methods in the strip footing example: (a) Octree- and traditional nonlocal computation methods; (b) proposed nonlocal model and local model.

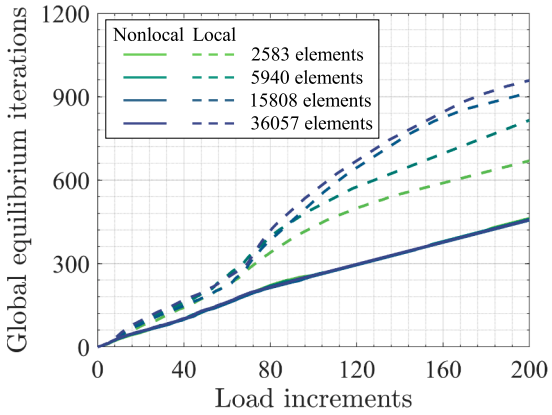


Fig. 23. Accumulated number of global equilibrium iterations for the proposed scaled nonlocal model and local model.

applicability. Future work will therefore prioritize the generalization to tensors and higher-dimensional proofs to unlock the method's full potential for advanced softening models and complex geotechnical problems. To facilitate reproducibility and further research, the UMAT subroutine associated with this work will be made open-source on GitHub (https://github.com/zhouxin615/scaled-nonlocal-model).

CRediT authorship contribution statement

Xin Zhou: Writing – original draft, Validation, Methodology, Funding acquisition. **Dechun Lu:** Writing – review & editing, Funding acquisition, Conceptualization. **Jidong Zhao:** Writing – review & editing, Supervision, Funding acquisition. **Yaning Zhang:** Visualization, Investigation, Formal analysis. **Zhiwei Gao:** Software, Resources, Methodology. **Timon Rabczuk:** Writing – review & editing, Funding acquisition. **Xiuli Du:** Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by the National Science Foundation of China (Grant Nos., 52025084, 52438005, 52408374, W2431038, 11972030) and Research Grants Council of Hong Kong (GRF 16212724, CRF C7082-22G, TRS T22-607/24N and T22-606/23-R).

Appendix A. . Derivation of the solution to the integral equation

Eq. (9) can be rewritten as:

$$\varepsilon^{\mathrm{p}}(x)+\frac{m}{2l_c(1-m)}\left[\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\right]=\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{A1}$$

The solution of Eq. (A1) depends on the value of m. When $m=1$, by simultaneously multiplying both sides of Eq. (A1) by $2l_c(1-m)$, one can obtain:

$$\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=2l_cH^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{A2}$$

Eq. (A2) is the Fredholm integral equation of the first kind, where the 1st derivatives of the two integrals with variable limits, $\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi$ and $\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi$, with respect to the variable x can be expressed as follows:

$$\begin{cases}\dfrac{\mathrm{d}\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi}{\mathrm{d}x}=\varepsilon^{\mathrm{p}}(x)-\dfrac{1}{l_c}\displaystyle\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\\ \dfrac{\mathrm{d}\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi}{\mathrm{d}x}=-\varepsilon^{\mathrm{p}}(x)+\dfrac{1}{l_c}\displaystyle\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\end{cases} \tag{A3}$$

Based on Eq. (A3), the 1st derivative of Eq. (A2) with respect to x can be obtained by:

$$\varepsilon^{\mathrm{p}}(x)-\frac{1}{l_c}\int_{-w/2}^{x} \mathrm{e}^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi-\varepsilon^{\mathrm{p}}(x)+\frac{1}{l_c}\int_{x}^{w/2} \mathrm{e}^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=0 \tag{A4}$$

Eq. (A4) simplifies to:

$$\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi-\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=0 \tag{A5}$$

Differentiating Eq. (A5) with respect to x yields:

$$\varepsilon^{\mathrm{p}}(x)-\frac{1}{l_c}\int_{-w/2}^{x} \mathrm{e}^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\varepsilon^{\mathrm{p}}(x)-\frac{1}{l_c}\int_{x}^{w/2} \mathrm{e}^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=0 \tag{A6}$$

After performing some manipulations, the following result can be obtained:

$$\varepsilon^{\mathrm{p}}(x)=\frac{1}{2l_c}\left[\int_{-w/2}^{x} \mathrm{e}^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\int_{x}^{w/2} \mathrm{e}^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\right] \tag{A7}$$

Substituting Eq. (A2) into Eq. (A7) yields the solution for $\varepsilon^{\mathrm{p}}(x)$ when $m=1$, as follows:

$$\varepsilon^{\mathrm{p}}(x)=H^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{A8}$$

It should be noted that, for $m=1$, the resulting $\varepsilon^{\mathrm{p}}(x)$ is a uniform solution, which is not suitable for the localization analysis. Next, we proceed to derive the solutions for $\varepsilon^{\mathrm{p}}(x)$ under the conditions of $m<1$ and $m>1$.

Differentiating Eq. (A1) with respect to x yields:

$$\frac{\partial\varepsilon^{\mathrm{p}}(x)}{\partial x}+\frac{m}{2l_c^2(1-m)}\left[-\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\right]=0 \tag{A9}$$

Then, differentiating Eq. (A9) once more with respect to x and applying Eq. (A3) leads to:

$$\frac{\partial^2\varepsilon^{\mathrm{p}}(x)}{\partial x^2}+\frac{m}{2l_c^2(1-m)}\left[\frac{1}{l_c}\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi-\varepsilon^{\mathrm{p}}(x)+\frac{1}{l_c}\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi-\varepsilon^{\mathrm{p}}(x)\right]=0 \tag{A10}$$

Eq. (A10) can be rewritten more concisely as follows:

$$\frac{\partial^2\varepsilon^{\mathrm{p}}(x)}{\partial x^2}-\frac{m\varepsilon^{\mathrm{p}}(x)}{l_c^2(1-m)}+\frac{m}{2l_c^3(1-m)}\left[\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\right]=0 \tag{A11}$$

To eliminate the integral term in Eq. (A11), we first perform a simple transformation on Eq. (A1):

$$\frac{m}{2l_c^3(1-m)}\left[\int_{-w/2}^{x} e^{\frac{\xi-x}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi+\int_{x}^{w/2} e^{\frac{x-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi\right]=\frac{1}{l_c^2}\left[\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)}-\varepsilon^{\mathrm{p}}(x)\right] \tag{A12}$$

Then, submitting Eq. (A12) into Eq. (A11) yields:

$$\frac{\partial^2\varepsilon^{\mathrm{p}}(x)}{\partial x^2}-\frac{m\varepsilon^{\mathrm{p}}(x)}{l_c^2(1-m)}+\frac{1}{l_c^2}\left[\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)}-\varepsilon^{\mathrm{p}}(x)\right]=0 \tag{A13}$$

Furthermore, Eq. (A13) can be rewritten as the following second-order non-homogeneous ordinary differential equation:

$$\frac{\partial^2 \varepsilon^{\mathrm{p}}(x)}{\partial x^2}+\frac{\varepsilon^{\mathrm{p}}(x)}{l_c^2(m-1)}=\frac{1}{l_c^2}\frac{H^{-1}(\sigma/\sigma_0)}{(m-1)} \tag{A14}$$

The characteristic equation, also referred to as the auxiliary equation, for Eq. (A14) is:

$$r^2+\frac{1}{l_c^2(m-1)}=0 \tag{A15}$$

For the case where $m<1$, Eq. (A15) has two different real roots:

$$r_1=\frac{1}{l_c\sqrt{1-m}}, r_2=-\frac{1}{l_c\sqrt{1-m}} \tag{A16}$$

Then, the general solution of the second-order homogeneous ordinary differential equation corresponding to Eq. (A14) is obtained as:

$$\varepsilon^{\mathrm{p}}(x)=C_1 \mathrm{e}^{\frac{x}{l_c\sqrt{1-m}}}+C_2 \mathrm{e}^{-\frac{x}{l_c\sqrt{1-m}}} \tag{A17}$$

where C_1 and C_2 are arbitrary constants.

For convenience in subsequent analysis, Eq. (A17) may be rewritten in terms of hyperbolic functions:

$$\varepsilon^{\mathrm{p}}(x)=C_1\left[\cosh\left(\frac{x}{l_c\sqrt{1-m}}\right)+\sinh\left(\frac{x}{l_c\sqrt{1-m}}\right)\right]+C_2\left[\cosh\left(\frac{x}{l_c\sqrt{1-m}}\right)-\sinh\left(\frac{x}{l_c\sqrt{1-m}}\right)\right] \tag{A18}$$

where $\cosh(x)=\frac{\mathrm{e}^x+\mathrm{e}^{-x}}{2}$ and $\sinh(x)=\frac{\mathrm{e}^x-\mathrm{e}^{-x}}{2}$ are the hyperbolic cosine function and the hyperbolic sine function, respectively. Eq. (A18) simplifies to:

$$\varepsilon^{\mathrm{p}}(x)=C\cosh\left(\frac{x}{l_c\sqrt{1-m}}\right)+D\sinh\left(\frac{x}{l_c\sqrt{1-m}}\right) \tag{A19}$$

where $C=C_1+C_2$ and $D=C_1-C_2$.

For $m>1$, Eq. (A15) has a pair of complex conjugate roots:

$$r_1=\frac{i}{l_c\sqrt{m-1}}, r_2=-\frac{i}{l_c\sqrt{m-1}} \tag{A20}$$

where i represents the imaginary unit.

Then, the general solution of the second-order homogeneous ordinary differential equation corresponding to Eq. (A14) for $m>1$ is given by:

$$\varepsilon^{\mathrm{p}}(x)=C\cos\left(\frac{x}{l_c\sqrt{m-1}}\right)+D\sin\left(\frac{x}{l_c\sqrt{m-1}}\right) \tag{A21}$$

It can be observed that $\varepsilon^{\mathrm{p}*}(x)=H^{-1}(\sigma/\sigma_0)$ is a specific solution of Eq. (A14).

$$\varepsilon^{\mathrm{p}*}(x)=H^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{A22}$$

Combing Eq. (A19), Eq. (A21), and Eq. (A22), we can yield the general solutions of Eq. (A14) under the conditions of $m<1$ and $m>1$:

$$\varepsilon^{\mathrm{p}}(x)=\begin{cases} C\cosh\left(\frac{x}{l_c\sqrt{1-m}}\right)+D\sinh\left(\frac{x}{l_c\sqrt{1-m}}\right)+H^{-1}\left(\frac{\sigma}{\sigma_0}\right) & m<1 \\ C\cos\left(\frac{x}{l_c\sqrt{m-1}}\right)+D\sin\left(\frac{x}{l_c\sqrt{m-1}}\right)+H^{-1}\left(\frac{\sigma}{\sigma_0}\right) & m>1 \end{cases} \tag{A23}$$

Due to the symmetry of the localization phenomenon, only the symmetric part of Eq. (A23) is reserved.

$$\varepsilon^{\mathrm{p}}(x)=\begin{cases} C\cosh\left(\frac{x}{l_c\sqrt{1-m}}\right)+H^{-1}\left(\frac{\sigma}{\sigma_0}\right) & m<1 \\ C\cos\left(\frac{x}{l_c\sqrt{m-1}}\right)+H^{-1}\left(\frac{\sigma}{\sigma_0}\right) & m>1 \end{cases} \tag{A24}$$

The function $\varepsilon^{\mathrm{p}}(x)$ satisfies the following Dirichlet and Neumann boundary conditions:

$$\varepsilon^{\mathrm{p}}\left(\pm\frac{w}{2}\right)=0, \left.\frac{\partial\varepsilon^{\mathrm{p}}(x)}{\partial x}\right|_{x=\pm\frac{w}{2}}=\pm\frac{1}{l_c}\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{A25}$$

where Eq. $(A25)_1$ denote that at the edge of the localization zone ($x=\pm w/2$), the plastic strain $\varepsilon^{\mathrm{p}}(x)$ must be equal to zero. The boundary conditions presented in Eq. $(A25)_2$ are nontrivial, derived from the properties of the Fredholm integral equation of the second kind, as discussed in the literature (Polyanin and Manzhirov, 2008). Appendix C presents a detailed derivation of Eq. $(A25)_2$. In this section, we directly present the results to maintain the readability of the paper.

First, we discuss the case where $m<1$. Substituting Eq. $(A24)_1$ into Eq. (A25) yields:

$$\varepsilon^{p}\left(\pm\frac{w}{2}\right)=C\cosh\left(\frac{\pm w}{2l_c\sqrt{1-m}}\right)+H^{-1}\left(\frac{\sigma}{\sigma_0}\right)=0 \tag{A26}$$

$$\left.\frac{\partial\varepsilon^{p}(x)}{\partial x}\right|_{x=\pm\frac{w}{2}}=\frac{C}{l_c\sqrt{1-m}}\sinh\left(\frac{\pm w}{2l_c\sqrt{1-m}}\right)=\pm\frac{1}{l_c}\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{A27}$$

Because $\varepsilon^{p}(x)$ is an even function and $\partial\varepsilon^{p}(x)/\partial x$ is an odd function, after some manipulations, Eq. (A26) and Eq. (A27) simplify to:

$$\cosh\left(\frac{w}{2l_c\sqrt{1-m}}\right)=-\frac{H^{-1}(\sigma/\sigma_0)}{C} \tag{A28}$$

$$\sinh\left(\frac{w}{2l_c\sqrt{1-m}}\right)=\frac{1}{C}\frac{H^{-1}(\sigma/\sigma_0)}{\sqrt{1-m}} \tag{A29}$$

Eq. (A29) divided by Eq. (A28) gives:

$$\tanh\left(\frac{w}{2l_c\sqrt{1-m}}\right)=-\frac{1}{\sqrt{1-m}}<0 \text{ and } m<1\Rightarrow w<0 \tag{A30}$$

Eq. (A30) means that the solution of $\varepsilon^{p}(x)$ has no physical meaning due to $w<0$. Next, we consider the case where $m>1$. Substituting Eq. $(A24)_2$ into Eq. (A25) yields:

$$\varepsilon^{p}\left(\pm\frac{w}{2}\right)=C\cos\left(\frac{\pm w}{2l_c\sqrt{m-1}}\right)+H^{-1}\left(\frac{\sigma}{\sigma_0}\right)=0 \tag{A31}$$

$$\left.\frac{\partial\varepsilon^{p}(x)}{\partial x}\right|_{x=\pm\frac{w}{2}}=-\frac{C}{l_c\sqrt{m-1}}\sin\left(\frac{\pm w}{2l_c\sqrt{m-1}}\right)=\pm\frac{1}{l_c}\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{A32}$$

Considering the Odevity of trigonometric functions, Eq. (A31) and Eq. (A32) simplify to:

$$\cos\left(\frac{w}{2l_c\sqrt{m-1}}\right)=-\frac{H^{-1}(\sigma/\sigma_0)}{C} \tag{A33}$$

$$\sin\left(\frac{w}{2l_c\sqrt{m-1}}\right)=\frac{H^{-1}(\sigma/\sigma_0)}{C\sqrt{m-1}} \tag{A34}$$

By combining Eq. (A33) and Eq. (A34) and using the trigonometric identity $\cos^2(x)+\sin^2(x)=1$, we obtain:

$$\cos^2\left(\frac{w}{2l_c\sqrt{m-1}}\right)+\sin^2\left(\frac{w}{2l_c\sqrt{m-1}}\right)=\frac{1}{C^2}\left[H^{-1}\left(\frac{\sigma}{\sigma_0}\right)\right]^2\left(\frac{m}{m-1}\right)=1 \tag{A35}$$

The expression for the integral constant C is obtained by solving Eq. (A35):

$$C=\sqrt{\frac{m}{m-1}}H^{-1}\left(\frac{\sigma}{\sigma_0}\right) \tag{A36}$$

Dividing Eq. (A34) by Eq. (A33) yields:

$$\tan\left(\frac{w}{2l_c\sqrt{m-1}}\right)=-\frac{1}{\sqrt{m-1}} \tag{A37}$$

The expression for the localization zone width w is obtained by solving Eq. (A37):

$$w=2l_c\sqrt{m-1}\left[n\pi-\arctan\left(\frac{1}{\sqrt{m-1}}\right)\right] \tag{A38}$$

where $n=1,\ 2,\ 3,\ldots$ denote the wave number. As the strain localization belongs to the case of short wave length, we take $n=1$ (Zhao et al., 2005). The localization zone width w is expressed as follows:

$$w=2l_c\sqrt{m-1}\left[\pi-\arctan\left(\frac{1}{\sqrt{m-1}}\right)\right] \tag{A39}$$

As $\arctan\left(1/\sqrt{m-1}\right)$ is a monotonically decreasing function with respect to m, we have:

$$\arctan\left(\frac{1}{\sqrt{m-1}}\right)<\lim_{m\to1^{+}}\arctan\left(\frac{1}{\sqrt{m-1}}\right)=\frac{\pi}{2} \tag{A40}$$

Substituting Eq. (A40) into Eq. (A39) yields:

$$w>l_c\pi\sqrt{m-1}>0, m>1 \tag{A41}$$

Eq. (A41) shows that the solution for the localization zone width w obtained under the condition of $m>1$ has physical meaning. Substituting Eq. (A36) into Eq. $(A24)_2$ yields the solution for the plastic strain $\varepsilon^{p}(x)$ within the localization zone.

$$\varepsilon^{p}(x)=H^{-1}\left(\frac{\sigma}{\sigma_0}\right)\left[\sqrt{\frac{m}{m-1}}\cos\left(\frac{x}{l_c\sqrt{m-1}}\right)+1\right],\quad -\frac{w}{2}\le x\le\frac{w}{2} \tag{A42}$$

It is important to note that the derivation above imposes no restrictions on the form of the softening function, implying its validity for any softening law.

Appendix B. . Shape functions of yield function and plastic potential function

The plasticity damage model presented in Section 3.1 adopts a modified Mohr-Coulomb yield function and non-associated flow rule in its plastic component. The standard Mohr-Coulomb function is often expressed in the form of principal stresses ($\overline{\sigma}_1 \geq \overline{\sigma}_2 \geq \overline{\sigma}_3$) as follows:

$$f = \frac{1}{2}(\overline{\sigma}_1 - \overline{\sigma}_3) + \frac{1}{2}(\overline{\sigma}_1 + \overline{\sigma}_3)\sin\varphi - c\cos\varphi = 0 \tag{B1}$$

where the effective stress is used since the plastic deformation is assumed to occur in the undamaged configuration of the material.

In numerical analysis of boundary value problems, the Mohr-Coulomb yield function expressed in the form of invariants ($\overline{p}$, $\overline{q}$, and $\overline{\theta}$) may be more common, as it avoids the tedious transformation between principal stress space and general stress space. The maximum principal stress $\overline{\sigma}_1$ and the minimum principal stress $\overline{\sigma}_3$ in Eq. (B1) can be expressed by:

$$\overline{\sigma}_1 = \overline{p} + \frac{2}{3}q\cos\overline{\theta}, \overline{\sigma}_3 = \overline{p} + \frac{2}{3}q\cos\left(\overline{\theta} + \frac{2\pi}{3}\right) \tag{B2}$$

Substituting Eq. (B2) into Eq. (B1) yields:

$$f = \frac{1}{3}\overline{q}\cos\overline{\theta} - \frac{1}{3}\overline{q}\cos\left(\overline{\theta} + \frac{2\pi}{3}\right) + \overline{p}\sin\varphi + \frac{1}{3}\overline{q}\cos\overline{\theta}\sin\varphi + \frac{1}{3}\overline{q}\cos\left(\overline{\theta} + \frac{2\pi}{3}\right)\sin\varphi - c\cos\varphi = 0 \tag{B3}$$

Performing some simple manipulations yields:

$$f = R_{MC}(\overline{\theta})\overline{q} + M\overline{p} - K = 0 \tag{B4}$$

where the expressions for $\overline{M}$ and $\overline{K}$ have been given in Eq. (32) and Eq. (33). $R_{\mathrm{MC}}(\overline{\theta})$ is the shape function of the Mohr-Coulomb criterion on the deviatoric plane, expressed as:

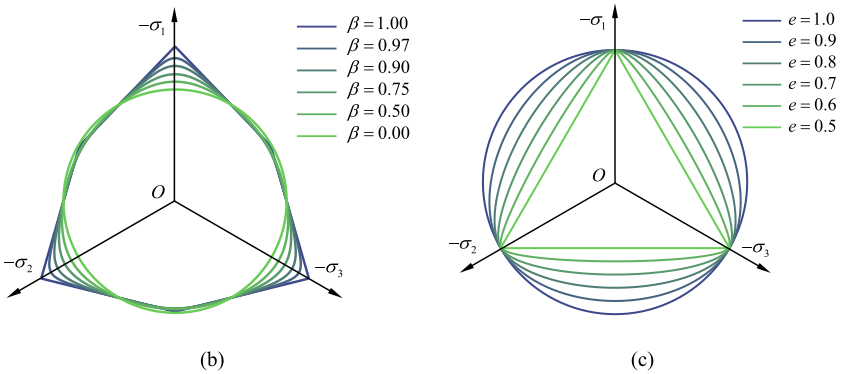
$$R_{MC}(\overline{\theta}) = \frac{2}{(3 - \sin\varphi)}\sin\left(\overline{\theta} + \frac{\pi}{3}\right) + \frac{2\sin\varphi}{\sqrt{3}(3 - \sin\varphi)}\cos\left(\overline{\theta} + \frac{\pi}{3}\right) \tag{B5}$$

The Mohr-Coulomb yield curve defined by Eq. (B5) is contained at six corner points on the deviatoric plane, which means that the derivative of the yield function in the vicinity of the corner points rapidly changes, potentially causing numerical instability. To this end, the smooth function $R_{mc}(\overline{\theta})$ proposed by Zhang et al. (2021) is used to replace Eq. (B5). The expression for $R_{mc}(\overline{\theta})$ is:

$$R_{mc} = \frac{\sqrt{3}}{3}\alpha_w\cos\left[\frac{1}{3}\arccos(\beta\cos 3\overline{\theta}) - \frac{\pi}{6}\gamma\right] \tag{B6}$$

where $\beta \in [0, 1]$ is a smoothing parameter. Fig. 24(a) illustrates the influence of β on the curves of shape function. As β increases, the shape of the yield curve gradually changes from a circle to an irregular hexagon. Eq. (B6) will degenerate into Eq. (B5) in the condition of $\beta = 1$. In this work, we set the default value $\beta = 0.999$ to ensure that the modified Mohr-Coulomb yield function remains sufficiently close to the standard criterion while eliminating the non-smoothness associated with the corner points. Parameters α_w and γ are functions of the internal friction angle:

$$\alpha_w = \frac{1}{\cos\left[(\overline{\gamma}_w + 1)\frac{\pi}{6}\right]}, \gamma = 1 - \overline{\gamma}_w, \overline{\gamma}_w = \frac{6}{\pi}\arctan\left(\frac{\sin\varphi}{\sqrt{3}}\right) \tag{B7}$$



(b) (c)

Fig. 24. Shape functions on the deviatoric plane: (a) yield curves; (b) plastic potential curves.

The deviatoric curve of the plastic potential function adopts the smooth elliptic function proposed by Menetrey and Willam (1995), which was also recommended by the ABAQUS software embedded Mohr-Coulomb model. The shape function of plastic potential function on the deviatoric plane is:

$$R_{mw} = \frac{4(1 - e^2)\cos^2\overline{\theta} + (2e - 1)^2}{2(1 - e^2)\cos\overline{\theta} + (2e - 1)\sqrt{4(1 - e^2)\cos^2\overline{\theta} + 5e^2 - 4e}}R_{mc}\left(\frac{\pi}{3}, \varphi\right) \tag{B8}$$

where e represents the eccentricity parameter. Fig. 24(b) illustrates the influence of e on the plastic potential shape curve. As e increases, the plastic potential curve gradually transitions from triangular to circular shape. $R_{mc}(\pi/3,\varphi)$ is the value of R_{mc} in the condition of $\overline{\theta}=\pi/3$ and $\beta=1$. The expressions for $R_{mc}(\pi/3,\varphi)$ and e are given as follows:

$$R_{mc}\left(\frac{\pi}{3},\varphi\right)=\frac{1}{2\cos\varphi}-\frac{1}{6}\tan\varphi, e=\frac{3-\sin\varphi}{3+\sin\varphi} \tag{B9}$$

Due to the intricate mathematical structure of the modified Mohr-Coulomb yield function and Menétrey-Willam plastic potential function, rigorously proving the thermodynamic admissibility of this model is highly challenging. Nonetheless, the conclusions regarding the thermodynamic admissibility of the standard Mohr-Coulomb model can serve as useful references. Hobbs and Ord (2014) analytically derived the rate of plastic work for non-associated Mohr-Coulomb materials during deformation (refer to pages 165 to 170 of the literature). Their findings indicate that the thermodynamic admissibility of the Mohr-Coulomb model depends on the material parameter values. Specifically, when $\psi\le\varphi$, the rate of the plastic work in Mohr-Coulomb materials is always non-negative, making it thermodynamically admissible. However, when $\psi>\varphi$, the rate of plastic work can become negative under certain conditions, potentially rendering Mohr-Coulomb materials thermodynamically inadmissible. In this case, the thermodynamic admissibility of the model depends on the current stress state and cohesive strength. Therefore, it is recommended to ensure $\psi\le\varphi$ for the Mohr-Coulomb model, which aligns with the deformation behaviour of geomaterials. Although the parameter range for thermodynamic admissibility in the standard Mohr-Coulomb model may not directly apply to the modified Mohr-Coulomb model discussed in this paper, the insights gained from the former still offer valuable guidance for parameter selection in the latter.

Appendix C. . Neumann boundary conditions of Eq. (A25)

By setting $x=-w/2$ and $x=w/2$ in Eq. (A1), two results can be obtained:

$$\varepsilon^{\mathrm{p}}\left(-\frac{w}{2}\right)+\frac{me^{-\frac{w}{2l_c}}}{2l_c(1-m)}\int_{-w/2}^{w/2}e^{\frac{-\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{C1}$$

$$\varepsilon^{\mathrm{p}}\left(\frac{w}{2}\right)+\frac{me^{-\frac{w}{2l_c}}}{2l_c(1-m)}\int_{-w/2}^{w/2}e^{\frac{\xi}{l_c}}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)} \tag{C2}$$

For simplicity in the derivation, we use $\lambda=-1/l_c$, $G=\frac{m}{2l_c(1-m)}$, $F=\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)}$, $a=-w/2$, and $b=w/2$. Then, Eq. (C1), Eq. (C2), and Eq. (A13)can be rewritten as follows:

$$\varepsilon^{\mathrm{p}}(a)+G\mathrm{e}^{-\lambda a}\int_a^b e^{\lambda\xi}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=F \tag{C3}$$

$$\varepsilon^{\mathrm{p}}(b)+G\mathrm{e}^{\lambda b}\int_a^b e^{-\lambda\xi}\varepsilon^{\mathrm{p}}(\xi)\mathrm{d}\xi=F \tag{C4}$$

$$\frac{\partial^2\varepsilon^{\mathrm{p}}(x)}{\partial x^2}+2\lambda G\varepsilon^{\mathrm{p}}(x)+\lambda^2[F-\varepsilon^{\mathrm{p}}(x)]=0 \tag{C5}$$

From Eq. (C5), we can express $\varepsilon^{\mathrm{p}}(\xi)$ as:

$$\varepsilon^{\mathrm{p}}(\xi)=\frac{1}{2G\lambda}\left[\lambda^2\varepsilon^{\mathrm{p}}(\xi)-\lambda^2F-\frac{\partial^2\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi^2}\right] \tag{C6}$$

Submitting Eq. (C6) into Eq. (C3) and performing some manipulations yields:

$$\int_a^b\lambda^2\varepsilon^{\mathrm{p}}(\xi)e^{\lambda\xi}\mathrm{d}\xi-\int_a^b\lambda^2Fe^{\lambda\xi}\mathrm{d}\xi-\int_a^b\frac{\partial^2\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi^2}e^{\lambda\xi}\mathrm{d}\xi=2\lambda\mathrm{e}^{\lambda a}[F-\varepsilon^{\mathrm{p}}(a)] \tag{C7}$$

Performing integration by parts on $\int_a^b\frac{\partial^2\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi^2}e^{\lambda\xi}\mathrm{d}\xi$, as shown in Eq. (C7), yields:

$$\int_a^b\frac{\partial^2\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi^2}e^{\lambda\xi}\mathrm{d}\xi=e^{\lambda b}\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=b}-e^{\lambda a}\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=a}-\lambda\left[\varepsilon^{\mathrm{p}}(b)e^{\lambda b}-\varepsilon^{\mathrm{p}}(a)e^{\lambda a}-\lambda\int_a^b\varepsilon^{\mathrm{p}}(\xi)e^{\lambda\xi}\mathrm{d}\xi\right] \tag{C8}$$

Submitting Eq. (C8) into Eq. (C7) yields:

$$e^{\lambda b}\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=b}-e^{\lambda a}\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=a}=\lambda\mathrm{e}^{\lambda a}[\varepsilon^{\mathrm{p}}(a)-F]+\lambda e^{\lambda b}[\varepsilon^{\mathrm{p}}(b)-F] \tag{C9}$$

Similarly, submitting Eq. (C6) into Eq. (C4) and simplifying gives:

$$e^{-\lambda b}\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=b}-e^{-\lambda a}\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=a}=\lambda\mathrm{e}^{-\lambda a}[\varepsilon^{\mathrm{p}}(a)-F]+\lambda e^{-\lambda b}[\varepsilon^{\mathrm{p}}(b)-F] \tag{C10}$$

Furthermore, by combing Eq. (C9) and Eq. (C10) and performing some manipulations, this gives the following result:

$$\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=a}=\lambda[F-\varepsilon^{\mathrm{p}}(a)],\frac{\partial\varepsilon^{\mathrm{p}}(\xi)}{\partial\xi}\bigg|_{\xi=b}=\lambda[\varepsilon^{\mathrm{p}}(b)-F] \tag{C11}$$

Substituting $\lambda=-1/l_c$, $F=\frac{H^{-1}(\sigma/\sigma_0)}{(1-m)}$, $a=-w/2$, $b=w/2$, and $\varepsilon^{\mathrm{p}}(\pm w/2)=0$ into Eq. (C11), we obtain the desired boundary conditions:

$$\left.\frac{\partial \varepsilon^{\mathrm{p}}(\xi)}{\partial \xi}\right|_{\xi=-\frac{w}{2}} = -\frac{H^{-1}(\sigma/\sigma_0)}{l_c(1-m)}, \left.\frac{\partial \varepsilon^{\mathrm{p}}(\xi)}{\partial \xi}\right|_{\xi=\frac{w}{2}} = \frac{H^{-1}(\sigma/\sigma_0)}{l_c(1-m)} \tag{C12}$$

It is evident that Eq. (C12) is equivalent to Eq. $(A25)_2$. A more general but somewhat less detailed derivation can be found in Section 4.9.37 of the literature (Polyanin and Manzhirov, 2008).

Data availability

Data will be made available on request.

References

Amirrahmat, S., Druckrey, A.M., Alshibli, K.A., Al-Raoush, R.I., 2019. Micro shear bands: Precursor for strain localization in sheared granular materials. J. Geotech. Geoenviron. Eng. 145 (2), 4018104. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001989.

Anand, L., Aslan, O., Chester, S.A., 2012. A large-deformation gradient theory for elastic–plastic materials: strain softening and regularization of shear bands. Int. J. Plast 30, 116–143. https://doi.org/10.1016/j.ijplas.2011.10.002.

Bažant, Z.P., Jirásek, M., 2002. Nonlocal integral formulations of plasticity and damage: survey of progress. J. Eng. Mech. 128 (11), 1119–1149. https://doi.org/10.1061/(ASCE)0733-9399(2002)128:11(1119).

Bažant, Z.P., Pijaudier-Cabot, G., 1989. Measurement of characteristic length of nonlocal continuum. J. Eng. Mech. 115 (4), 755–767. https://doi.org/10.1061/(ASCE)0733-9399(1989)115:4(755).

Bédorf, J., Gaburov, E., Zwart, S.P., 2012. A sparse octree gravitational N-body code that runs entirely on the GPU processor. J. Comput. Phys. 231 (7), 2825–2839. https://doi.org/10.1016/j.jcp.2011.12.024.

Behley, J., Steinhage, V., Cremers, A.B., 2015. *Efficient radius neighbor search in three-dimensional point clouds*. In: 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 3625–3630.

Challamel, N., 2008. A regularization study of some softening beam problems with an implicit gradient plasticity model. J. Eng. Math. 62, 373–387. https://doi.org/10.1007/s10665-008-9233-3.

Challamel, N., Lanos, C., Casandjian, C., 2008. Plastic failure of nonlocal beams. Phys. Rev. E—Statis., Nonlinear, Soft Matter Phys. 78 (2), 26604. https://doi.org/10.1103/PhysRevE.78.026604.

Chen, P.C., Feng, Y., Ma, Y.E., 2025. A high-accuracy phase field model for complex cohesive law. Eng. Fract. Mech. 315, 110765. https://doi.org/10.1016/j.engfracmech.2024.110765.

Chen, J., Hawlader, B., Roy, K., Pike, K., 2023. A nonlocal Eulerian-based finite-element approach for strain-softening materials. Comput. Geotech. 154, 105114. https://doi.org/10.1016/j.compgeo.2022.105114.

Cui, W.J., Wu, X.T., Potts, D.M., Zdravković, L., 2023. Nonlocal strain regularisation for critical state models with volumetric hardening. Comput. Geotech. 157, 105350. https://doi.org/10.1016/j.compgeo.2023.105350.

de Berg, M., Cheong, O., Van Kreveld, M., Overmars, M., 2008. Computational Geometry: Algorithms and Applications. Springer, Berlin Heidelberg, Berlin, Heidelberg.

de Borst, R., 1987. Computation of post-bifurcation and post-failure behavior of strain-softening solids. Comput. Struct. 25 (2), 211–224. https://doi.org/10.1016/0045-7949(87)90144-1.

de Borst, R., 1991. Simulation of strain localization: a reappraisal of the Cosserat continuum. Eng. Comput. 8 (4), 317–332. https://doi.org/10.1108/eb023842.

de Borst, R., Verhoosel, C.V., 2016. Gradient damage vs phase-field approaches for fracture: similarities and differences. Comput. Methods Appl. Mech. Eng. 312, 78–94. https://doi.org/10.1016/j.cma.2016.05.015.

Denarie, E., Saouma, V.E., Iocco, A., Varelas, D., 2001. Concrete fracture process zone characterization with fiber optics. J. Eng. Mech. 127 (5), 494–502. https://doi.org/10.1061/(ASCE)0733-9399(2001)127:5(494).

Di Luzio, G., Bažant, Z.P., 2005. Spectral analysis of localization in nonlocal and over-nonlocal materials with softening plasticity or damage. Int. J. Solids Struct. 42 (23), 6071–6100. https://doi.org/10.1016/j.ijsolstr.2005.03.038.

Dorduncu, M., Ren, H.L., Zhuang, X.Y., Silling, S., Madenci, E., Rabczuk, T., 2024. A review of peridynamic theory and nonlocal operators along with their computer implementations. Comput. Struct. 299, 107395. https://doi.org/10.1016/j.compstruc.2024.107395.

Engelen, R.A., Geers, M.G., Baaijens, F.P., 2003. Nonlocal implicit gradient-enhanced elasto-plasticity for the modelling of softening behaviour. Int. J. Plast 19 (4), 403–433. https://doi.org/10.1016/S0749-6419(01)00042-0.

Feng, Y., Hai, L., 2025. 3D phase-field cohesive fracture: Unifying energy, driving force, and stress criteria for crack nucleation and propagation direction. J. Mech. Phys. Solids. 196, 106036. https://doi.org/10.1016/j.jmps.2025.106036.

Fernández-Fernández, J.A., Westhofen, L., Löschner, F., Jeske, S.R., Longva, A., Bender, J., 2022. Fast octree neighborhood search for SPH simulations. ACM Trans. Graphics (TOG) 41 (6), 1–13. https://doi.org/10.1145/3550454.3555523.

Finno, R.J., Harris, W.W., Mooney, M.A., Viggiani, G., 1997. Shear bands in plane strain compression of loose sand. Geotechnique 47 (1), 149–165. https://doi.org/10.1680/geot.1997.47.1.149.

Galavi, V., Schweiger, H.F., 2010. Nonlocal multilaminate model for strain softening analysis. Int. J. Geomech. 10 (1), 30–44. https://doi.org/10.1061/(ASCE)1532-3641(2010)10:1(30).

Grassl, P., Jirásek, M., 2006. Plastic model with non-local damage applied to concrete. Int. J. Numer. Anal. Meth. Geomech. 30 (1), 71–90. https://doi.org/10.1002/nag.479.

Hazeghian, M., Soroush, A., 2022. DEM-aided study of Coulomb and Roscoe theories for shear band inclination. Acta Geotech. 17 (8), 3357–3375. https://doi.org/10.1007/s11440-022-01475-y.

Hobbs, B.E., Ord, A., 2014. Structural Geology: The Mechanics of Deforming Metamorphic Rocks Principles. Elsevier, Amsterdam.

Huang, M.S., Qu, X., Lü, X.L., 2018. Regularized finite element modelling of progressive failure in soils within nonlocal softening plasticity. Comput. Mech. 62 (3), 347–358. https://doi.org/10.1007/s00466-017-1500-6.

Iacono, C., Sluys, L.J., van Mier, J.G., 2006. Estimation of model parameters in nonlocal damage theories by inverse analysis techniques. Comput. Methods Appl. Mech. Eng. 195 (52), 7211–7222. https://doi.org/10.1016/j.cma.2004.12.033.

Jirásek, M., 2007. Nonlocal damage mechanics. Revue européenne de génie civil 11 (7–8), 993–1021. https://doi.org/10.1080/17747120.2007.9692974.

Jirásek, M., Rolshoven, S., Grassl, P., 2004. Size effect on fracture energy induced by non-locality. Int. J. Numer. Anal. Meth. Geomech. 28 (7–8), 653–670. https://doi.org/10.1002/nag.364.

Karapiperis, K., Ortiz, M., Andrade, J.E., 2021. Data-driven nonlocal mechanics: discovering the internal length scales of materials. Comput. Methods Appl. Mech. Eng. 386, 114039. https://doi.org/10.1016/j.cma.2021.114039.

Le Bellégo, C., Dubé, J.F., Pijaudier-Cabot, G., Gérard, B., 2003. Calibration of nonlocal damage model from size effect tests. Eur. J. Mechanics-A/Solids 22 (1), 33–46. https://doi.org/10.1016/S0997-7538(02)01255-X.

Lee, J., Fenves, G.L., 1998. Plastic-damage model for cyclic loading of concrete structures. J. Eng. Mech. 124 (8), 892–900. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:8(892).

Li, X., Gao, Z.W., 2024. Evaluation of three weight functions for nonlocal regularization of sand models. Int. J. Geomech. 24 (7), 4024125. https://doi.org/10.1061/IJGNAI.GMENG-9192.

Lu, X.L., Bardet, J., Huang, M.S., 2009. Numerical solutions of strain localization with nonlocal softening plasticity. Comput. Methods Appl. Mech. Eng. 198 (47–48), 3702–3711. https://doi.org/10.1016/j.cma.2009.08.002.

Lü, X.L., Zhao, Y.C., Xue, D.W., Lim, K.W., Qin, H.L., 2023. Numerical modelling of shield tunnel face failure through a critical state sand plasticity model with nonlocal regularization. Comput. Geotech. 164, 105847. https://doi.org/10.1016/j.compgeo.2023.105847.

Mallikarachchi, H., Soga, K., 2020. Post-localisation analysis of drained and undrained dense sand with a nonlocal critical state model. Comput. Geotech. 124, 103572. https://doi.org/10.1016/j.compgeo.2020.103572.

Mánica, M.A., Gens Solé, A., Vaunat, J., Ruiz, D.F., 2017. *Analysis of strain localization with a nonlocal plasticity model*. In: COMPLAS XIV: proceedings of the XIV International Conference on Computational Plasticity: fundamentals and applications. CIMNE, pp. 606–612.

Mánica, M.A., Gens, A., Vaunat, J., Ruiz, D.F., 2018. Nonlocal plasticity modelling of strain localisation in stiff clays. Comput. Geotech. 103, 138–150. https://doi.org/10.1016/j.compgeo.2018.07.008.

Marcher, T. 2003. *Nichtlokale Modellierung der Entfestigung dichter Sande und steifer Tone*. Ph.D. Thesis, University of Stuttgart, Stuttgart, Germany.

Mazzucco, G., Xotta, G., Salomoni, V.A., Majorana, C., 2019. Integral-type regularization of non associated softening plasticity for quasi brittle materials. Comput. Struct. 224, 106120. https://doi.org/10.1016/j.compstruc.2019.106120.

Menetrey, P., Willam, K.J., 1995. Triaxial failure criterion for concrete and its generalization. Struct. J. 92 (3), 311–318. https://doi.org/10.14359/1132.

Miehe, C., Hofacker, M., Welschinger, F., 2010. A phase field model for rate-independent crack propagation: Robust algorithmic implementation based on operator splits. Comput. Methods Appl. Mech. Eng. 199 (45–48), 2765–2778. https://doi.org/10.1016/j.cma.2010.04.011.

Mohamed, A., Gutierrez, M., 2010. Comprehensive study of the effects of rolling resistance on the stress–strain and strain localization behaviour of granular materials. Granul. Matter 12, 527–541. https://doi.org/10.1007/s10035-010-0211-x.

Moreau, K., Moës, N., Picart, D., Stainier, L., 2015. Explicit dynamics with a non-local damage model using the thick level set approach. Int. J. Numer. Meth. Eng. 102 (3–4), 808–838. https://doi.org/10.1002/nme.4824.

Otsuka, K., Date, H., 2000. Fracture process zone in concrete tension specimen. Eng. Fract. Mech. 65 (2–3), 111–131. https://doi.org/10.1016/S0013-7944(99)00111-3.

Polyanin, P., Manzhirov, A.V., 2008. Handbook of Integral Equations. Chapman and Hall/CRC.

Ram, P., Sinha, K., 2019. *Revisiting kd-tree for nearest neighbor search*. In: Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining, pp. 1378–1388.

Rattez, H., Shi, Y., Sac-Morane, A., Klaeyle, T., Mielniczuk, B., Veveakis, M., 2022. Effect of grain size distribution on the shear band thickness evolution in sand. Géotechnique 72 (4), 350–363. https://doi.org/10.1680/jgeot.20.P.120.

Rechenmacher, A.L., 2006. Grain-scale processes governing shear band initiation and evolution in sands. J. Mech. Phys. Solids 54 (1), 22–45. https://doi.org/10.1016/j.jmps.2005.08.009.

Ren, H.L., Zhuang, X.Y., Rabczuk, T., 2020. A nonlocal operator method for solving partial differential equations. Comput. Methods Appl. Mech. Eng. 358, 112621. https://doi.org/10.1016/j.cma.2019.112621.

Rolshoven, S. 2003. *Nonlocal plasticity models for localized failure*. Ph.D. Thesis, Swiss Federal Institute of Technology at Lausanne (EPFL).

Samoladas, D., Karras, C., Karras, A., Theodorakopoulos, L., Sioutas, S., 2022. *Tree data structures and efficient indexing techniques for big data management: a comprehensive study*. In: Proceedings of the 26th Pan-Hellenic Conference on Informatics, pp. 123–132.

Scherzinger, W.M., 2017. A return mapping algorithm for isotropic and anisotropic plasticity models using a line search method. Comput. Methods Appl. Mech. Eng. 317, 526–553. https://doi.org/10.1016/j.cma.2016.11.026.

Shi, J.S., Guan, L., Zhuang, D.Y., Chen, X., Ling, D.S., 2023. Fault rupture propagation in soil with intercalation using nonlocal model and softening modulus modification. J. Rock Mech. Geotech. Eng. 15 (11), 2973–2993. https://doi.org/10.1016/j.jrmge.2023.06.003.

Silling, S.A., Lehoucq, R.B., 2010. Peridynamic theory of solid mechanics. Adv. Appl. Mech. 44, 73–168. https://doi.org/10.1016/S0065-2156(10)44002-8.

Singh, V., Stanier, S., Bienen, B., Randolph, M.F., 2021. Modelling the behaviour of sensitive clays experiencing large deformations using non-local regularisation techniques. Comput. Geotech. 133, 104025. https://doi.org/10.1016/j.compgeo.2021.104025.

Su, C.C., Lu, D.C., Zhou, X., Wang, G.S., Zhuang, X.Y., Du, X.L., 2023. An implicit stress update algorithm for the plastic nonlocal damage model of concrete. Comput. Methods Appl. Mech. Eng. 414, 116189. https://doi.org/10.1016/j.cma.2023.116189.

Sun, Z.C., Chu, J., Xiao, Y., 2021. Formulation and implementation of an elastoplastic constitutive model for sand-fines mixtures. Int. J. Numer. Anal. Meth. Geomech. 45 (18), 2682–2708. https://doi.org/10.1002/nag.3282.

Taiebat, M., Dafalias, Y.F., 2008. SANISAND: simple anisotropic sand plasticity model. Int. J. Numer. Anal. Meth. Geomech. 32 (8), 915–948. https://doi.org/10.1002/nag.651.

Tang, J.J., Kong, X.Z., Fang, Q., Hong, J., Fan, J.Y., 2021. An efficient three-dimensional damage-based nonlocal model for dynamic tensile failure in concrete. Int. J. Impact Eng 156, 103965. https://doi.org/10.1016/j.ijimpeng.2021.103965.

Varas, F., Alonso, E., Alejano, L.R., Manín, G.F., 2005. Study of bifurcation in the problem of unloading a circular excavation in a strain-softening material. Tunn. Undergr. Space Technol. 20 (4), 311–322. https://doi.org/10.1016/j.tust.2004.12.003.

Vazic, B., Diyaroglu, C., Oterkus, E., Oterkus, S., 2020. Family member search algorithms for peridynamic analysis. J. Peridynamics Nonlocal Model. 2 (1), 59–84. https://doi.org/10.1007/s42102-019-00027-5.

Vermeer, P.A., 1990. The orientation of shear bands in biaxial tests. Géotechnique 40 (2), 223–236. https://doi.org/10.1680/geot.1990.40.2.223.

Wu, J.Y., 2017. A unified phase-field theory for the mechanics of damage and quasi-brittle failure. J. Mech. Phys. Solids 103, 72–99. https://doi.org/10.1016/j.jmps.2017.03.015.

Wu, W.L., Zhao, L.Y., Lai, Y.M., Lv, Z.M., Chen, Y.Y., Ran, J.C., 2024. A rational multiscale nonlinear constitutive model for freeze–thaw rocks under triaxial compression. Int. J. Plasticity. 179, 104040. https://doi.org/10.1016/j.ijplas.2024.104040.

Xenos, D., Grégoire, D., Morel, S., Grassl, P., 2015. Calibration of nonlocal models for tensile fracture in quasi-brittle heterogeneous materials. J. Mech. Phys. Solids 82, 48–60. https://doi.org/10.1016/j.jmps.2015.05.019.

Yang, C.Y., Zhu, F., Zhao, J.D., 2024. A multi-horizon fully coupled thermo-mechanical peridynamics. J. Mech. Phys. Solids 191, 105758. https://doi.org/10.1016/j.jmps.2024.105758.

Yu, H.T., Hu, X.K., Bobet, A., Yan, X., 2024. A nonlocal hybrid model for elasto-plastic fracture of rock-like materials. Comput. Methods Appl. Mech. Eng. 424, 116884. https://doi.org/10.1016/j.cma.2024.116884.

Yu, H.T., Sun, Y.Q., 2021. Bridging the gap between local and nonlocal numerical methods—a unified variational framework for non-ordinary state-based peridynamics. Comput. Methods Appl. Mech. Eng. 384, 113962. https://doi.org/10.1016/j.cma.2021.113962.

Yu, Z.P., Yang, Y., Zhang, J., Peng, S., 2024. Research on uniaxial mechanical performance of high-performance concrete after high temperature rapid cooling and damage mechanism analysis. J. Build. Eng. 86, 108921.

Zhang, N., Li, X., Wang, D., 2021. Smoothed classic yield function for C2 continuities in tensile cutoff, compressive cap, and deviatoric sections. Int. J. Geomech. 21 (3), 4021005. https://doi.org/10.1061/(ASCE)GM.1943-5622.0001910.

Zhao, J.D., Sheng, D.C., Zhou, W.Y., 2005. Shear banding analysis of geomaterials by strain gradient enhanced damage model. Int. J. Solids Struct. 42 (20), 5335–5355. https://doi.org/10.1016/j.ijsolstr.2005.02.029.

Zhou, X., Lu, D.C., Du, X.L., Wang, G.S., Meng, F.P., 2020. A 3D non-orthogonal plastic damage model for concrete. Comput. Methods Appl. Mech. Eng. 360, 112716. https://doi.org/10.1016/j.cma.2019.112716.

Zhou, X., Lu, D.C., Su, C.C., Gao, Z.W., Du, X.L., 2022. An unconstrained stress updating algorithm with the line search method for elastoplastic soil models. Comput. Geotech. 143, 104592. https://doi.org/10.1016/j.compgeo.2021.104592.

Zhou, X., Shi, A.Y., Lu, D.C., Chen, Y., Zhuang, X.Y., Lu, X.Z., Du, X.L., 2023. A return mapping algorithm based on the hyper dual step derivative approximation for elastoplastic models. Comput. Methods Appl. Mech. Eng. 417, 116418. https://doi.org/10.1016/j.cma.2023.116418.

Zhuang, L., Nakata, Y., Kim, U., Kim, D., 2014. Influence of relative density, particle shape, and stress path on the plane strain compression behavior of granular materials. Acta Geotech. 9 (2), 241–255. https://doi.org/10.1007/s11440-013-0253-4.