Understanding Nonlinear Preconditioning Techniques for Newton's Method from a Dynamical Systems Perspective

Hwang, Feng-Nan[1]
[1]National Central University

1 Introduction

Nonlinear preconditioning is a powerful tool for accelerating the convergence of Newton-type methods applied to large, sparse, and unbalanced nonlinear systems [1]. Classical globalization strategies such as line search or trust-region techniques often struggle in stiff systems, where rapid variations in some components constrain step size, much like in stiff ODEs. This paper adopts a dynamical systems perspective, interpreting Newton's method as a continuous-time flow (the "Newton ODE") and introducing a stiffness ratio to quantify nonlinearity imbalance—analogous to the condition number in linear systems [4]. This metric offers insight into convergence behavior and may help guide the design of effective nonlinear preconditioners. We illustrate this idea through a model problem and numerical experiments using nonlinear elimination preconditioned inexact Newton methods (NEPIN) [3].

2 Newton ODE and stiffness ratio

The Newton iteration with backtracking for solving a nonlinear system $F(x) = 0$ is given by $x^{(k+1)} = x^{(k)} - \lambda J^{-1}(x^{(k)})F(x^{(k)})$, where $J(x)$ is the Jacobian and $\lambda \in (0, 1]$ is the step size. In the continuous-time limit, this becomes the Newton ODE: $J(x)\frac{dx}{dt} = -F(x)$. Linearizing around a steady state x^* with $F(x^*) = 0$ yields: $B\frac{dr}{dt} = Ar$, where $r = x - x^*$, $A = -J(x^*)$, and $B = J(x)$. The dynamics are governed by the generalized eigenvalue problem $Av = \lambda Bv$, and we define the stiffness ratio as

$$S = \frac{\max_j |Re(\lambda_j)|}{\min_j |Re(\lambda_j)|}.$$

A large S indicates stiffness, motivating the use of nonlinear preconditioning. To address this, we consider the following two Newton algorithms:

- **INB** a classical globalization approach that uses line search or trust-region strategies for robustness.
- **NEPIN+HF (NEPIN with Hole Filling)** [2]: a left-preconditioned method that solves a nonlinear subproblem on selected variables and applies a morphological operation to improve component selection.

In NEPIN+HF, the system is partitioned into "good" and "bad" components. A coarse-grid correction and morphological closing operation are applied to improve accuracy in identifying stiff regions. This preconditioner has demonstrated improved performance for problems with complex wave structures.

3 Application to the Buckley–Leverett equation

We consider the 1D BL equation $u_t + [f(u)]_x = 0$, with the flux function $f(u) = \frac{u^2}{u^2+a(1-u)^2}$, where a is a constant representing the water-to-oil viscosity ratio. The variable u represents water saturation, ranging from 0 (pure oil) to 1 (pure water).

4 Stiffness Analysis for INB and NEPIN

We examine the stiffness behavior of the Newton ODE systems arising from the BL equation, focusing on two canonical cases: a single-shock scenario and a two-phase flow configuration. In

 Copyright (C) 2025 一般社団法人日本応用数理学会

both cases, the initial guess is constructed by interpolating a coarse-grid solution on a 8×4 grid. The simulations are performed on a 256×128 grid.

Figure 1 presents the evolution of the stiffness ratio over Newton iterations for both INB and NEPIN+HF methods. Across all cases, the stiffness ratio decreases monotonically and eventually converges to 1. For this problem, which exhibits stronger nonlinear imbalance, the stiffness ratio is reduced by more than 50% relative to INB. This early reduction in stiffness effectively stabilizes the system and enhances convergence speed for the preconditioned Newton solver.

5 Conclusion

We introduced a stiffness-based perspective on nonlinear preconditioning by interpreting Newton's method as an ODE system. The stiffness ratio serves as a useful metric for identifying imbalance in nonlinear systems. Numerical experiments confirm that preconditioners such as NEPIN+HF reduce stiffness early in the iteration, improving convergence.

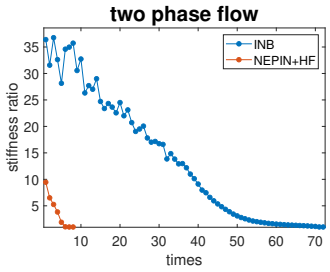

Figure 1: Evolution of stiffness ratio for INB and NEPIN+HF in selected test cases.

References

[1] F.-N. Hwang, Y.-C. Su, and X.-C. Cai, A parallel adaptive nonlinear elimination preconditioned inexact Newton method for transonic full potential equation, Computers and Fluids, Vol. 110 (2015) pp. 96–107.

[2] C.-W. Liang and F.-N. Hwang, Nonlinear elimination preconditioned space-time solution algorithms for hyperbolic PDE problems, submitted, 2024.

[3] L. Liu, F.-N. Hwang, L. Luo, X.-C. Cai, D.E. Keyes, Nonlinear elimination preconditioned inexact Newton algorithm, SIAM Journal on Scientific Computing, Vol. 44, (2022), pp. A1579–A1605.

[4] L. Luo, C.-W. Liang, F.-N. Hwang, and X.-C. Cai, Insight into nonlinear preconditioning from a dynamical system viewpoint, in preparation, May 2025.

 Copyright (C) 2025 一般社団法人日本応用数理学会

Generating rough paths and simulating diffusions: algorithms and machine learning implementations

Toshihiro Yamada[1]
[1]Hitotsubashi University
e-mail : toshihiro.yamada@r.hit-u.ac.jp

1 Rough path-based algorithm for simulating diffusions

On a Wiener space $(\Omega, \mathcal{F}, P)$ where $\Omega := C_{(0)}([0,T];\mathbb{R}^d) := \{\omega \in C([0,T];\mathbb{R}^d); \omega(0) = 0\}$, $\mathcal{F} := \mathscr{B}(\Omega)$ and P is the Wiener measure, consider a diffusion process given by the following Stratonovich SDE:

$$dX_t^x = \sum_{i=0}^d V_i(X_t^x) \circ dB_t^i, \quad X_0^x = x \in \mathbb{R}^N. \tag{1}$$

where $\circ dB_t^0 = dt$ and $V_i \in C_b^\infty(\mathbb{R}^N;\mathbb{R}^N)$, $i = 0, 1, \ldots, d$. Let $B_{s,t}^i := B_s^i - B_t^i$, $s > t$, $i = 0, 1, \ldots, d$. Roughly speaking, the rough path theory tells us that the Itô-Lyons map $\mathbf{B} \mapsto X_\cdot^x$ is continuous where the notation $\mathbf{B}$ represents "Brownian rough path" $\mathbf{B} = ((B^i)_i, (\mathbb{B}^{(i,j)})_{i,j})$ with $\mathbb{B}^{(i,j)} = \int_{0<t_1<t_2<\cdot} \circ dB_{t_1}^i \circ dB_{t_2}^j$, $1 \le i, j \le d$ (Friz and Hairer [1]). The following rough path-based scheme of Naito and Yamada [2] provides a better approximation than the Euler-Maruyama scheme in both weak and strong sense, which is ensured by the theorem below.

Algorithm 1 Rough path-based scheme

1: ——— Approximation of rough path ———
2: **for** $k = 1$ to n **do**
3: **for** $i, j = 1$ to d **do**
4: $\bar{\mathbb{B}}^{(i),n}_{t_k,t_{k-1}} = \sum_{\ell=1}^n B^i_{\tau_\ell^{k-1},\tau_{\ell-1}^{k-1}}$, $\bar{\mathbb{B}}^{(i,j),n}_{t_k,t_{k-1}} = \sum_{\ell=1}^n B^i_{\tau_{\ell-1}^{k-1},t_{k-1}} B^j_{\tau_\ell^{k-1},\tau_{\ell-1}^{k-1}}$ with $\tau_\ell^{k-1} = t_{k-1} + \ell T/n^2$, $\ell = 1, \ldots, n$
5: **end for**
6: **end for**
7: ——— Approximation of SDE solution path ———
8: **for** $k = 1$ to n **do**
9: $x^{\mathrm{RP},n}_{t_k \leftarrow 0} = x^{\mathrm{RP},n}_{t_{k-1} \leftarrow 0}$
$+ \sum_{\alpha \sim_e \beta, e \in \mathbb{N}, |\beta| \le 2}^{\|\alpha\| \le 4,} V_{\alpha_1} \cdots V_{\alpha_{|\alpha|}}(x^{\mathrm{RP},n}_{t_{k-1} \leftarrow 0}) \sum_{\alpha \sim_e \gamma, e \in \mathbb{N}, |\gamma| \le 2} \frac{1}{2^{\nu(\alpha,\gamma)}} \bar{\mathbb{B}}^{\gamma,n}_{t_k,t_{k-1}}$
10: **end for**

Theorem 1 (Naito and Yamada [2])**.**

1) (*Weak approximation*) *It holds that*

$$\left|\mathbb{E}[f(X_T^x)] - \mathbb{E}[f(x^{\mathrm{RP},n}_{t_n \leftarrow 0}(x))]\right| = O(n^{-2}) \tag{2}$$

for all $f \in C_b^\infty(\mathbb{R}^N)$, whose rate is better than that of the Euler-Maruyama scheme (Kloeden and Platen [3]): $\left|\mathbb{E}[f(X_T^x)] - \mathbb{E}[f(x^{\mathrm{EM}}_{t_n \leftarrow 0}(x))]\right| = O(n^{-1})$ *for all $f \in C_b^\infty(\mathbb{R}^N)$.*

2) (*Strong approximation*) *For $p \ge 2$, it holds that*

$$\mathbb{E}[|X_T^x - x^{\mathrm{RP},n}_{t_n \leftarrow 0}(x)|^p]^{\frac{1}{p}} = O(n^{-1}), \tag{3}$$

whose rate is better than that of the Euler-Maruyama scheme (Maruyama [4]): $\mathbb{E}[|X_T^x - x^{\mathrm{EM}}_{t_n \leftarrow 0}(x)|^p]^{\frac{1}{p}} = O(n^{-1/2})$.

One can extend the above algorithm to higher order cases, in other words, for each $m \ge 2$, a rough path scheme $\{x^{\mathrm{RP},(m),n}_{t_i \leftarrow 0}(x)\}_{i \le n}$ with approximate iterated Stratonovich integrals

 Copyright (C) 2025 一般社団法人日本応用数理学会

$(x^{\mathrm{RP},(2),n}_{t_i\leftarrow 0}(x) = x^{\mathrm{RP},n}_{t_i\leftarrow 0}(x))$ is constructed in a similar manner, in order to obtain more accurate weak and strong approximations.

2 Applications

The rough path-based scheme is useful for simulating diffusions and provides wide applications in applied mathematics. For example, consider the following semilinear PDE:

$$(\partial_t + \mathcal{L})u(t,x) + \ell(x, u(t,x)) = 0, \quad u(T,\cdot) = f \tag{4}$$

with $f \in C_b^\infty(\mathbb{R}^N), \ell \in C_b^\infty(\mathbb{R}^N \times \mathbb{R})$ and $\mathcal{L} = \hat{V}_0 + (1/2)\sum_{i=1}^d \hat{V}_i^2$ where $\hat{V}_i = \sum_{j=1}^N V_i^j(\cdot)\frac{\partial}{\partial x_j}$. In financial mathematics and theory of backward stochastic differential equations, there is interest in the problem of solving the PDE at $t = 0$ and $x \in \mathbb{R}^N$, i.e. $u(0,x)$. Based on the rough path-based scheme, it can be efficiently solved by a high order discretization algorithm with machine learning (deep learning) technique, even if the dimension N is high. For a fixed $m \geq 2$, we recursively define functions

$$\begin{aligned} u^{(m)}(t_i,\cdot) = \mathrm{argmin}_{v\in L^2(\mathbb{R}^N,\mathscr{B}(\mathbb{R}^N),P\circ(x^{\mathrm{RP},(m),n}_{t_i\leftarrow 0}(x))^{-1})}\mathbb{E}\big[\big|v(x^{\mathrm{RP},(m),n}_{t_i\leftarrow 0}(x)) \\ - \textstyle\sum_{k=1}^m a_k u^{(m)}(t_{(i+k)\wedge n}, x^{\mathrm{RP},(m),n}_{t_{(i+k)\wedge n}\leftarrow 0}(x)) \\ - \frac{T}{n}\textstyle\sum_{k=0}^m b_k \ell(x^{\mathrm{RP},(m),n}_{t_{(i+k)\wedge n}\leftarrow 0}(x), u^{(m)}(t_{(i+k)\wedge n}, x^{\mathrm{RP},(m),n}_{t_{(i+k)\wedge n}\leftarrow 0}(x)))\big|^2\big], \quad i = n-1,\dots,1,0, \end{aligned} \tag{5}$$

where $t_i = iT/n$, $i = 0,1,\dots,n$, and $\{a_j\}_{j=1,\dots,m}$, $\{b_j\}_{j=0,1,\dots,m}$ satisfy that $\sum_{j=1}^m a_j = 1$, $a_j \geq 0$, $1 \leq j \leq m$ and $\sum_{j=1}^m a_j j^p - p\sum_{j=0}^m j^{p-1}b_j = 0$, $1 \leq p \leq m$. Here, for $i = 0$, it should be understood as the L^2-minimization representation for $u^{(m)}(0,x) = \mathbb{E}[\sum_{k=1}^m a_k u^{(m)}(t_{k\wedge n}, x^{\mathrm{RP},(m),n}_{t_{k\wedge n}\leftarrow 0}(x)) + \frac{T}{n}\sum_{k=0}^m b_k \ell(x^{\mathrm{RP},(m),n}_{t_{k\wedge n}\leftarrow 0}(x), u^{(m)}(t_{k\wedge n}, x^{\mathrm{RP},(m),n}_{t_{k\wedge n}\leftarrow 0}(x)))]$. Then, it holds that $|u(0,x) - u^{(m)}(0,x)| = O(n^{-m})$.

Acknowledgements

I would like to thank Prof. Riu Naito (University of Toyama) for his many valuable comments on the content of this talk.

References

[1] P. Friz and M. Hairer, *A Course on Rough Paths*, Springer (2014)

[2] R. Naito, and T. Yamada, A rough path-based scheme for weak and strong approximations of SDEs, preprint (submitted) (2025).

[3] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer (1992)

[4] G. Maruyama, Continuous Markov processes and stochastic equations, *Rend. Circ. Mat. Palermo*, 4, 48–90 (1955)

 Copyright (C) 2025 一般社団法人日本応用数理学会

Multimodal Mathematical Reasoning with Large Language Model

Yoonsuk Hyun[1]
[1]Inha University
e-mail : yshyun21@inha.ac.kr

1 Introduction

Multimodal mathematical reasoning (MMR) stands at the intersection of computer vision, natural language processing, and symbolic reasoning. Unlike traditional text-only math reasoning, MMR requires models to integrate diverse data modalities, such as diagrams, charts, or visual cues, with rigorous symbolic manipulation. This capability is crucial for solving real-world problems that inherently combine visual and formal mathematical elements.

2 Motivation

As highlighted in the slides and summarized in recent position papers, MMR covers a broad scope: diverse reasoning types, visual elements, and multimodal alignments. A major motivation is to develop AI systems that perform reliable perception and verifiable reasoning. Recent show how far AI can push the boundaries of competitive mathematics when formal systems complement informal reasoning. Expanding research in this direction toward multimodal mathematical reasoning that jointly leverages vision and language data is a natural progression.

3 Benchmarks and Datasets

High-quality benchmarks are vital for meaningful progress. A comprehensive benchmark spanning visual reasoning and symbolic math. These datasets highlight the diversity of tasks and the need for rigorous evaluation metrics. Effective benchmarking demands clear recipes: integrating multimodal inputs, domain knowledge, and rigorous reasoning. Evaluation should not rely solely on answer correctness but also on the quality of intermediate reasoning steps. Concepts like LLM-as-judge have been proposed for more nuanced assessment.

4 Techniques and Foundation Models

With the advent of techniques that effectively align image and language data for joint training, large language models (LLMs) have begun to expand into vision-language models. The introduction of CLIP[1], which uses image-text pairs through contrastive learning, marked a major breakthrough in this area. This progress has led to the development of powerful vision-language foundation models. A key factor behind their impressive performance is the precise alignment between visual and linguistic modalities.

In the field of mathematical reasoning, several foundation models have been developed, resulting in algorithms capable of solving a wide range of math problems with remarkable accuracy.

Math-LLaVA[2] uses 40K high-quality images and 320K synthesized image–question–answer pairs. This approach boosts multimodal mathematical reasoning performance, and shows

Proceedings of the 2025 Annual Meeting of the Japan Society for Industrial and Applied Mathematics
Copyright (C) 2025 The Japan Society for Industrial and Applied Mathematics

stronger generalizability, outperforming the base model and other open-source MLLMs. MAVIS[3] introduces a four-stage training pipeline designed to enhance multimodal LLMs' performance on visual math problems.

Math-PUMA[4] proposes a three-stage training framework to enhance mathematical reasoning in multimodal large language models. It first strengthens the model on text-only mathematical problems, then progressively aligns textual and visual modalities using KL divergence across a large paired dataset, and finally fine-tunes on nearly one million multimodal problems. This progressive alignment strategy significantly improves the model' s ability to handle vision-rich math problems and achieves state-of-the-art performance on several benchmarks.

GeoX[5] is a multimodal model tailored for geometry problem solving, introducing a three-stage formalized training pipeline—unimodal pre-training for diagram encoding and symbol decoding, geometry–language alignment using a novel Generator-and-Sampler Transformer (GS-Former), and end-to-end visual instruction tuning for verifiable solution generation. This approach bridges the modality gap between visual diagrams and formal symbolic language, enabling the model to produce structured, program-style solutions that are both interpretable and correct.

5 Challenges and Future Works

Major challenges in this area include the scarcity of high-quality, well-aligned visual-math data, as well as the need for robust benchmarks. Another key difficulty lies in developing architectures that effectively integrate visual perception, formal verification, and rigorous intermediate reasoning. Future research is expected to advance toward more robust perceptual capabilities, transparent reasoning processes.

References

[1] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PmLR.

[2] Shi, W., Hu, Z., Bin, Y., Liu, J., Yang, Y., Ng, S. K., ... & Lee, R. K. W. (2024). Math-llava: Bootstrapping mathematical reasoning for multimodal large language models. arXiv preprint arXiv:2406.17294.

[3] Zhang, R., Wei, X., Jiang, D., Zhang, Y., Guo, Z., Tong, C., ... & Li, H. (2024). Mavis: Mathematical visual instruction tuning. arXiv e-prints, arXiv-2407.

[4] Zhuang, W., Huang, X., Zhang, X., & Zeng, J. (2025, April). Math-puma: Progressive upward multimodal alignment to enhance mathematical reasoning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 24, pp. 26183-26191).

[5] Xia, R., Li, M., Ye, H., Wu, W., Zhou, H., Yuan, J., ... & Zhang, B. (2024). Geox: Geometric problem solving through unified formalized vision-language pre-training. arXiv preprint arXiv:2412.11863.

Copyright (C) 2025 The Japan Society for Industrial and Applied Mathematics