A Preconditioned Version of a Nested Primal-Dual Algorithm for Image Deblurring

Giuseppe Scarlato[1]

Joint work with:

Stefano Aleotti[1]

Marco Donatelli[1]

Rolf Krause[2]

July 29, 2025

[1] Università degli Studi dell'Insubria

[2] King Abdullah University of Science and Technology (KAUST)

Optimization problem

Consider the following optimization problem:

$$\underset{u\in\mathbb{R}^d}{\operatorname{argmin}} \; f(u) + h(Wu), \tag{1}$$

Optimization problem

Consider the following optimization problem:

$$\operatorname*{argmin}_{u \in \mathbb{R}^d} \underbrace{f(u)}_{\text{smooth}} + h(Wu), \tag{1}$$

where:

- $f : \mathbb{R}^d \to \mathbb{R}$ is a convex and smooth function,

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

Consider the following optimization problem:

$$\underset{u\in\mathbb{R}^d}{\operatorname{argmin}} \underbrace{f(u)}_{\text{smooth}} + \underbrace{h(Wu)}_{\text{convex}}, \tag{1}$$

where:

- $f : \mathbb{R}^d \to \mathbb{R}$ is a convex and smooth function,
- $h : \mathbb{R}^{d'} \to \mathbb{R} \cup \{\infty\}$ is convex and possibly non-smooth,
- $W \in \mathbb{R}^{d' \times d}$ is a linear operator.

Consider the following optimization problem:

$$\underset{u \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{f(u)}_{\text{smooth}} + \underbrace{h(Wu)}_{\text{convex}}, \tag{1}$$

where:

- $f : \mathbb{R}^d \to \mathbb{R}$ is a convex and smooth function,
- $h : \mathbb{R}^{d'} \to \mathbb{R} \cup \{\infty\}$ is convex and possibly non-smooth,
- $W \in \mathbb{R}^{d' \times d}$ is a linear operator.

Image reconstruction applications

- Image deblurring
- Denoising
- Computed Tomography

Proximal gradient

Proximal gradient methods are first-order methods that alternate a gradient step on f and a proximal evaluation on $h \circ W$:

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

Proximal gradient methods are first-order methods that alternate a gradient step on f and a proximal evaluation on $h \circ W$:

$$\begin{cases} u_{n+\frac{1}{2}} = u - \alpha_n \nabla f(u) & \leftarrow \text{gradient descent} \\ \end{cases}$$

Proximal gradient

Proximal gradient methods are first-order methods that alternate a gradient step on f and a proximal evaluation on $h \circ W$:

$$\begin{cases} u_{n+\frac{1}{2}} = u - \alpha_n \nabla f(u) & \leftarrow \text{gradient descent} \\ u_{n+1} = \operatorname{prox}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}) & \leftarrow \text{proximal evaluation} \end{cases} \tag{2}$$

Proximal gradient

Proximal gradient methods are first-order methods that alternate a gradient step on f and a proximal evaluation on $h \circ W$:

$$\begin{cases} u_{n+\frac{1}{2}} = u - \alpha_n \nabla f(u) & \leftarrow \text{gradient descent} \\ u_{n+1} = \operatorname{prox}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}) & \leftarrow \text{proximal evaluation} \end{cases} \tag{2}$$

where

- $\alpha_n > 0$ is the step size,
- $\operatorname{prox}_g(v) = \operatorname{argmin}_u \ \frac{1}{2}\|u - v\|^2 + g(u)$ is the proximal operator of g.

Proximal gradient

Proximal gradient methods are first-order methods that alternate a gradient step on f and a proximal evaluation on $h \circ W$:

$$\begin{cases} u_{n+\frac{1}{2}} = u - \alpha_n \nabla f(u) & \leftarrow \text{gradient descent} \\ u_{n+1} = \operatorname{prox}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}) & \leftarrow \text{proximal evaluation} \end{cases} \tag{2}$$

where

- $\alpha_n > 0$ is the step size,
- $\operatorname{prox}_g(v) = \operatorname{argmin}_u \ \frac{1}{2}\|u - v\|^2 + g(u)$ is the proximal operator of g.

A Nesterov-like extrapolation, like in FISTA, can be added to eq. (2) to accelerate convergence.

Proximal gradient methods are first-order methods that alternate a gradient step on f and a proximal evaluation on $h \circ W$:

$$\begin{cases} u_{n+\frac{1}{2}} = u - \alpha_n \nabla f(u) & \leftarrow \text{gradient descent} \\ u_{n+1} = \operatorname{prox}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}) & \leftarrow \text{proximal evaluation} \end{cases} \tag{2}$$

where

- $\alpha_n > 0$ is the step size,
- $\operatorname{prox}_g(v) = \operatorname{argmin}_u \ \frac{1}{2}\|u - v\|^2 + g(u)$ is the proximal operator of g.

A Nesterov-like extrapolation, like in FISTA, can be added to eq. (2) to accelerate convergence.

Equation (2) assumes that the proximal operator of $h \circ W$ can be computed in closed form. This is not the case for some common regularization terms, such as the Total Variation (TV).

The Nested Primal-Dual (NPD)[1,2] algorithm is an inexact proximal gradient method summarized as

$$\begin{cases} \bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}) & \leftarrow \text{Nesterov extrap.} \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n \nabla f(\bar{u}_n) & \leftarrow \text{gradient descent} \\ u_{n+1} \approx \operatorname{prox}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}) & \leftarrow \text{inexact proximal step} \end{cases}$$

where $\gamma_n \geq 0$ is the extrapolation parameter like in FISTA.

[1] Chen and Loris 2019, "On starting and stopping criteria for nested primal-dual iterations".
[2] Bonettini et al. 2023, "A nested primal–dual FISTA-like scheme for composite convex optimization problems".

Nested Primal-Dual (NPD)

The Nested Primal-Dual (NPD)[1,2] algorithm is an inexact proximal gradient method summarized as

$$\begin{cases} \bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}) & \leftarrow \text{Nesterov extrap.} \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n \nabla f(\bar{u}_n) & \leftarrow \text{gradient descent} \\ u_{n+1} \approx \text{prox}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}) & \leftarrow \text{inexact proximal step} \end{cases}$$

where $\gamma_n \geq 0$ is the extrapolation parameter like in FISTA.

The inexact proximal step is computed by $k_{max} \in \mathbb{N}$ steps of a dual sequence v^k:
Let $u \in \mathbb{R}^d$, $v^0 \in \mathbb{R}^{d'}$, $\alpha > 0$, $0 < \beta < 2/\|W\|^2$, and define the sequence

$$v^{k+1} = \text{prox}_{\beta\alpha^{-1}h^*}(v^k + \beta\alpha^{-1}W(u - \alpha W^T v^k)),$$

having limit $\lim_{k\to\infty} v^k = \hat{v}$. Then we have

$$\text{prox}_{\alpha h \circ W}(u) = u - \alpha W^T \hat{v},$$

[1] Chen and Loris 2019, "On starting and stopping criteria for nested primal-dual iterations".
[2] Bonettini et al. 2023, "A nested primal–dual FISTA-like scheme for composite convex optimization problems".

$$\operatorname*{argmin}_{u\in\mathbb{R}^d} \underbrace{\frac{1}{2}\|Au - b^\delta\|^2}_{\text{data fidelity}} + \underbrace{\lambda\,\mathrm{TV}(u)}_{\text{regularization}},$$

which is a specific instance of the model in eq. (1) where

- $f(u) = \frac{1}{2}\|Au - b^\delta\|^2$,
- $h(Wu) = \lambda\,\mathrm{TV}(u)$,
- $A \in \mathbb{R}^{d\times d}$ is the blurring operator,
- $\lambda > 0$ is a regularization parameter
- $b^\delta \in \mathbb{R}^d$ is the observed image: the blurred image with noise.

Image deblurring

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

$$\underset{u \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|Au - b^\delta\|^2}_{\text{data fidelity}} + \underbrace{\lambda \operatorname{TV}(u)}_{\text{regularization}},$$

which is a specific instance of the model in eq. (1) where

- $f(u) = \frac{1}{2}\|Au - b^\delta\|^2$,
- $h(Wu) = \lambda \operatorname{TV}(u)$,
- $A \in \mathbb{R}^{d \times d}$ is the blurring operator,
- $\lambda > 0$ is a regularization parameter
- $b^\delta \in \mathbb{R}^d$ is the observed image: the blurred image with noise.

The TV is the Total Variation operator defined as

$$\operatorname{TV}(x) = \sum_{i=1}^{d} \|\nabla_i u\|.$$

$W \in \mathbb{R}^{2d \times d}$ is the discretization of the gradient operator, and $h : \mathbb{R}^{2d} \to \mathbb{R}$ is the sum of the pixel-wise euclidean norms.

The Nested Primal-Dual Iterated Tikhonov (NPDIT)[3] is a variant of NPD which uses a variable metric approach to achieve faster convergence:

$$\begin{cases} \bar{u}_n = u_n + \gamma_n (u_n - u_{n-1}), \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n P^{-1} \nabla f(\bar{u}_n), \\ u_{n+1} \approx \mathrm{prox}^{P}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}). \end{cases}$$

[3] Aleotti et al. 2025, "A nested primal–dual iterated Tikhonov method for regularized convex optimization".

The Nested Primal-Dual Iterated Tikhonov (NPDIT)[3] is a variant of NPD which uses a variable metric approach to achieve faster convergence:

$$\begin{cases} \bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}), \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n P^{-1}\nabla f(\bar{u}_n), \\ u_{n+1} \approx \mathrm{prox}^{P}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}). \end{cases}$$

For linear inverse problems, a suitable choice for P_n is

$$P_n = A^T A + \nu_n I.$$

[3] Aleotti et al. 2025, "A nested primal–dual iterated Tikhonov method for regularized convex optimization".

Nested Primal-Dual Iterated Tikhonov (NPDIT)

The Nested Primal-Dual Iterated Tikhonov (NPDIT)[3] is a variant of NPD which uses a variable metric approach to achieve faster convergence:

$$\begin{cases} \bar{u}_n = u_n + \gamma_n (u_n - u_{n-1}), \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n P^{-1} \nabla f(\bar{u}_n), \\ u_{n+1} \approx \mathrm{prox}^{P}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}). \end{cases}$$

For linear inverse problems, a suitable choice for P_n is

$$P_n = A^T A + \nu_n I.$$

With this choice, $u_{n+\frac{1}{2}}$ is obtained as an Iterated Tikhonov, or, equivalently, as a Levenberg-Marquardt step on the data fidelity term.

[3] Aleotti et al. 2025, "A nested primal–dual iterated Tikhonov method for regularized convex optimization".

Preconditioning

Given the linear system

$$Au = b,$$

the preconditioner can be applied on both sides of A:

Given the linear system

$$Au = b,$$

the preconditioner can be applied on both sides of A:

Left preconditioning (PNPD)

$$P^{-1}Au = P^{-1}b$$

Given the linear system

$$Au = b,$$

the preconditioner can be applied on both sides of A:

Left preconditioning (PNPD)

$$P^{-1}Au = P^{-1}b$$

Right preconditioning (NPDIT)

$$AP^{-1}Pu = b$$

NPDIT as right preconditioned NPD

Applying the right preconditioning to the least squares regularized problem, we obtain the equivalent formulation

$$\hat{u} = \operatorname*{argmin}_{u \in \mathbb{R}^d} \frac{1}{2} \| A R^{-1} R u - b^\delta \|^2 + h(Wu). \tag{3}$$

NPDIT as right preconditioned NPD

Applying the right preconditioning to the least squares regularized problem, we obtain the equivalent formulation

$$\hat{u} = \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} \frac{1}{2} \| A R^{-1} R u - b^\delta \|^2 + h(Wu). \tag{3}$$

The NPD method applied to problem (3) results in

$$\begin{cases} \bar{u}_n = u_n + \gamma_n (u_n - u_{n-1}), \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n (R^T R)^{-1} \nabla f(\bar{u}_n), \\ u_{n+1} \approx \operatorname{prox}^{R^T R}_{\alpha_n h \circ W} (u_{n+\frac{1}{2}}), \end{cases}$$

Applying the right preconditioning to the least squares regularized problem, we obtain the equivalent formulation

$$\hat{u} = \operatorname*{argmin}_{u \in \mathbb{R}^d} \frac{1}{2}\|AR^{-1}Ru - b^\delta\|^2 + h(Wu). \tag{3}$$

The NPD method applied to problem (3) results in

$$\begin{cases} \bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}), \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n(R^T R)^{-1}\nabla f(\bar{u}_n), \\ u_{n+1} \approx \mathrm{prox}^{R^T R}_{\alpha_n h \circ W}(u_{n+\frac{1}{2}}), \end{cases}$$

which is NPDIT with the right preconditioner

$$P_n = R^T R.$$

Applying the left preconditioning to the least squares regularized problem, the norm of the data fidelity term changes. Given S positive definite, we have

$$\operatorname*{argmin}_{u \in \mathbb{R}^d} \frac{1}{2} \|Au - b^\delta\|^2_{S^{-1}} + h(Wu), \tag{4}$$

where

$$f_S(u) = \|Au - b^\delta\|^2_{S^{-1}} = \|S^{-\frac{1}{2}}(Au - b^\delta)\|^2.$$

Applying the left preconditioning to the least squares regularized problem, the norm of the data fidelity term changes. Given S positive definite, we have

$$\underset{u\in\mathbb{R}^d}{\operatorname{argmin}} \frac{1}{2}\|Au - b^\delta\|^2_{S^{-1}} + h(Wu), \tag{4}$$

where

$$f_S(u) = \|Au - b^\delta\|^2_{S^{-1}} = \|S^{-\frac{1}{2}}(Au - b^\delta)\|^2.$$

Assuming that there exist a positive definite matrix P such that

$$P^{-1}A^T = A^T S^{-1}, \tag{5}$$

Applying the left preconditioning to the least squares regularized problem, the norm of the data fidelity term changes. Given S positive definite, we have

$$\operatorname*{argmin}_{u\in\mathbb{R}^d} \frac{1}{2}\|Au - b^\delta\|^2_{S^{-1}} + h(Wu), \tag{4}$$

where

$$f_S(u) = \|Au - b^\delta\|^2_{S^{-1}} = \|S^{-\frac{1}{2}}(Au - b^\delta)\|^2.$$

Assuming that there exist a positive definite matrix P such that

$$P^{-1}A^T = A^T S^{-1}, \tag{5}$$

then NPD applied to eq. (4) gives the Preconditioned NPD (PNPD) algorithm:

$$\begin{cases} \bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}), \\ u_{n+\frac{1}{2}} = \bar{u}_n - \alpha_n P^{-1}\nabla f(\bar{u}_n), \\ u_{n+1} \approx \operatorname{prox}_{\alpha_n h\circ W}(u_{n+\frac{1}{2}}), \end{cases}$$

Preconditioner choice

Choosing the preconditioner P as in NPDIT, the condition in eq. (5) is satisfied, since

$$P = A^T A + \nu I, \qquad S = AA^T + \nu I,$$

with $\nu > 0$. In general, the identity in eq. (5) is satisfied whenever P is a polynomial of $A^T A$ and S is a corresponding polynomial of AA^T.

PNPD

NPDIT

PNPD

NPDIT

PNPD uses the standard definition of the proximity operator of $h \circ W$, while NPDIT uses a variable metric approach. Therefore, to compute the dual sequence, PNPD does not require multiplying by P^{-1} opposed to NPDIT.

PNPD

PNPD uses the standard definition of the proximity operator of $h \circ W$, while NPDIT uses a variable metric approach. Therefore, to compute the dual sequence, PNPD does not require multiplying by P^{-1} opposed to NPDIT.

NPDIT

The data fidelity of the optimization problem in eq. (4) uses a weighted norm defined by the positive definite matrix S instead of the usual Euclidean norm as in eq. (3)

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

Theorem (Convergence of PNPD)

Let $\{(u_n, v_n^0)\}_{n\in\mathbb{N}}$ be the primal–dual sequence generated by the PNPD method with $\alpha_n = \alpha \in \left(0, \frac{1}{L_S}\right]$, where L_S is the Lipschitz constant of ∇f_S, and $\beta_n = \beta \in (0, \|W\|^{-2})$ for all $n \in \mathbb{N}$. Suppose also that the extrapolation parameters $\{\gamma_n\}_{n\in\mathbb{N}}$ satisfies

$$\sum_{n=0}^{\infty} \gamma_n \|u_n - u_{n-1}\| < \infty,$$

and that S and P satisfy equation (5).

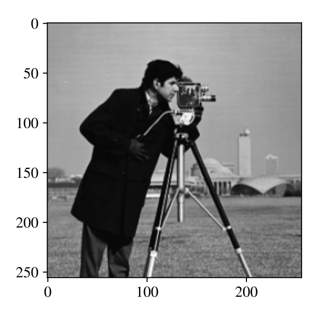
Then, the following statements hold:

(i) the sequence $\{(u_n, v_n^0)\}_{n\in\mathbb{N}}$ is bounded;

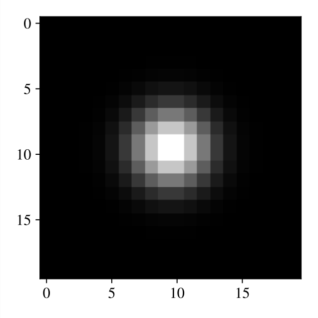
(ii) the primal sequence $\{u_n\}_{n\in\mathbb{N}}$ converges to a solution of the initial problem (4).

Numerical results

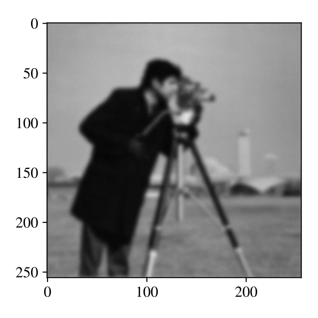
Image deblurring example



(a) Ground truth image



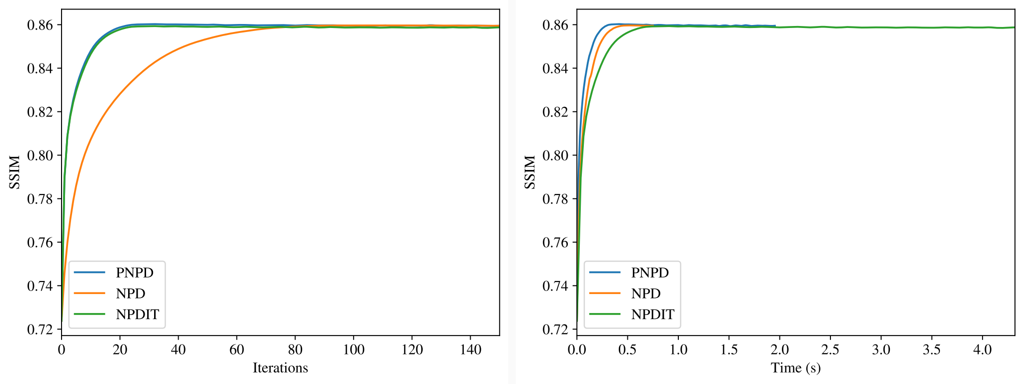
(b) Gaussian PSF



(c) Observed image b^δ

- The ground truth image is a 256×256 grayscale image of a cameramen.
- The PSF is a Gaussian with standard deviation $\sigma = 2$ pixels.
- The Gaussian noise η_δ is such that $\|\eta_\delta\| = 0.01\|b^\delta\|$.

Structural Similarity Index (SSIM) results



Parameters:

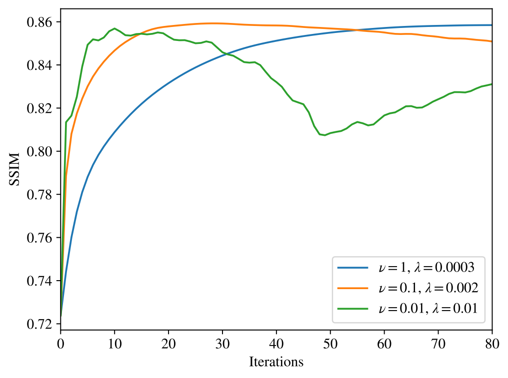
- Preconditioner parameter: $\nu = 10^{-1}$
- Regularization parameter: $\lambda = 2 \cdot 10^{-4}$ for NPD and NPDIT, while $\lambda = 2 \cdot 10^{-3}$ for PNPD.
- Number of nested loop iterations: $k_{\text{max}} = 1$ for NPD and $k_{\text{max}} = 3$ for NPDIT and PNPD.

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

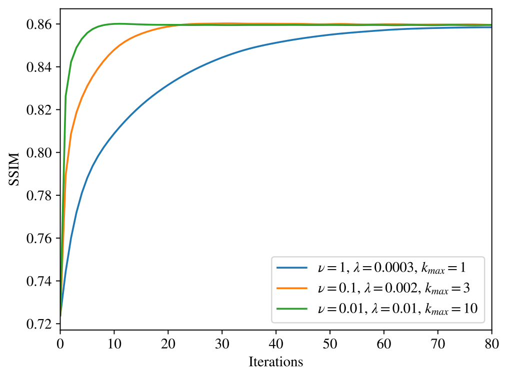
k_{max}	PNPD	NPDIT	Δ	NPDIT/PNPD
1	0.0107	0.0196	0.0088	1.822
2	0.0109	0.0245	0.0135	2.239
4	0.0125	0.0372	0.0247	2.971
8	0.0194	0.0617	0.0423	3.177
16	0.0239	0.0933	0.0694	3.894
32	0.0357	0.1508	0.1151	4.216
64	0.0584	0.2703	0.2119	4.629

Table 1: Average time spent for one step of PNPD and NPDIT for different values of $k_{\max}$. Δ is the difference between the execution time of the two methods.

PNPD as ν changes



(a) $k_{\max} = 1$ for different values of ν and λ.



(b) $k_{\max}$ set high enough to fix instability.

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

Conclusions

- The convergence of PNPD is guaranteed under suitable assumptions.
- PNPD converges to the solution of a problem with a energy norm instead of the usual Euclidean norm in the data fidelity term.
- Numerical results show that PNPD is more efficient than NPD and NPDIT in terms of CPU time, especially for large values of k_{max}.

Conclusions

- The convergence of PNPD is guaranteed under suitable assumptions.
- PNPD converges to the solution of a problem with a energy norm instead of the usual Euclidean norm in the data fidelity term.
- Numerical results show that PNPD is more efficient than NPD and NPDIT in terms of CPU time, especially for large values of k_{max}.

Future work

- Approximation of P^{-1}.
- Apply PNPD to other image reconstruction problems such as computed tomography.
- Unfolding of PNPD to learn optimal parameters.

References

S. Aleotti, M. Donatelli, R. Krause, G. Scarlato
A Preconditioned Version of a Nested Primal-Dual Algorithm for Image Deblurring
J. Sci. Comput. 103, 85 (2025)

Official GitHub repository for the PNPD codes
`https://github.com/Giuseppe499/PNPD`

Thank you for your attention!