

INDUSTRIAL AND SYSTEMS ENGINEERING



Stochastic Galerkin Method and Hierarchical Preconditioning for PDE-Constrained Optimization

ZHENDONG LI¹, AKWUM ONWUNTA¹, AND BEDŘICH SOUSEDÍK²

¹Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, 18015
USA

²Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000
Hilltop Circle, Baltimore, MD 21250 USA

ISE Technical Report 25T-024-R1



STOCHASTIC GALERKIN METHOD AND HIERARCHICAL PRECONDITIONING FOR PDE-CONSTRAINED OPTIMIZATION *

ZHENDONG LI[†], AKWUM ONWUNTA[†], AND BEDŘICH SOUSEDÍK[‡]

Abstract. We develop efficient hierarchical preconditioners for optimal control problems governed by partial differential equations with uncertain coefficients. Adopting a discretize-then-optimize framework that integrates finite element discretization, stochastic Galerkin projection, and advanced time-discretization schemes, the approach addresses challenges of scaling large and ill-conditioned linear systems arising in uncertainty quantification. By exploiting sparsity of linear systems in stochastic Galerkin method, we formulate hierarchical preconditioners based on truncated stochastic expansion that strike an effective balance between computational cost and preconditioning quality. Numerical experiments demonstrate that the proposed preconditioners significantly accelerate the convergence of iterative solvers compared to existing methods, providing robust and efficient solvers for both steady-state and time-dependent optimal control problems under uncertainty.

Key words. Stochastic Galerkin method, preconditioning, iterative solvers, Gauss-Seidel method, hierarchical and multilevel preconditioning

MSC codes. 35R60, 65C20, 65F08, 65F10, 60H35, 65N22

1. Introduction. Optimal control problems governed by partial differential equations (PDEs) arise in numerous applications, including fluid mechanics, structural optimization, and inverse problems. These problems have been extensively studied over the past decades. For a theoretical overview and computational methods related to deterministic problems, we refer readers to, e.g., [9, 27]. In many practical applications, the PDE coefficients are uncertain. Such uncertainties originate from various sources, including measurement errors, model approximations, and environmental variations, and they are modeled as random variables or stochastic processes. Recently, there has been an increased interest in optimal control problems governed by PDEs with random coefficients, see e.g., [17]. These stochastic problems are inherently more complex than their deterministic counterparts, thus necessitating specialized numerical methods.

Two strategies are commonly used: optimize-then-discretize and discretize-then-optimize. We adopt the latter, which discretizes the objective and PDE first, enabling the use of efficient numerical methods like finite elements [1, 3, 4, 5, 9].

For stochastic discretization, we employ the stochastic Galerkin method [11, 14, 15, 16, 22, 29], which expands the solution using orthonormal polynomials. Although the Monte Carlo method is simpler and the stochastic collocation method [8, 13] decouples the problem, stochastic Galerkin method systematically captures uncertainties and preserves optimal convergence properties, offering computational efficiency for large-scale problems when carefully implemented.

In order to address the PDE-constrained optimization problems, we employ the discretize-then-optimize approach. In practice, it is common to combine temporal discretization (e.g., backward Euler scheme), stochastic expansions (e.g., generalized polynomial chaos expansions), and spatial discretization (e.g., finite element method).

*Dedicated to the memory of Professor Howard C. Elman, our dear colleague and mentor. We also want to thank Lehigh's High Performance Computing systems for providing computational resources.

[†]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015 (zh1923@lehigh.edu, ako221@lehigh.edu).

[‡]Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250 (sousedik@umbc.edu).

This leads to large-scale linear systems obtained via finite element (or possibly finite difference) discretizations, which are then typically solved using Krylov subspace methods, e.g., by the generalized minimal residual method (GMRES). These linear systems are often ill-conditioned, which causes a slow convergence of the iterative method. To address this issue, we construct preconditioners that improve the convergence of iterative solvers.

In this paper, we introduce a hierarchical preconditioning framework specifically tailored for stochastic PDE-constrained optimal control problems. Although the core concepts are inspired by preconditioners for PDE problems [6, 25, 26, 30], an application to the Karush-Kuhn-Tucker (KKT) systems arising from optimization problems and an extension to time-dependent problems are nontrivial and constitute a primary contribution of this work. We provide a systematic derivation of the preconditioner for both steady-state and all-at-once formulation of time-dependent problems. The method is supported by a spectral analysis, proving that the proposed preconditioner is spectrally equivalent to the ideal but computationally prohibitive exact Schur complement. The performance is evaluated using a set of numerical experiments.

The paper is organized as follows. In Section 2.1, we introduce the steady-state stochastic optimal control problem and its discretization into a large-scale KKT system. Then, we construct the hierarchical Gauss-Seidel preconditioner for PDE-constrained optimal control problems (hGSoc). In Section 3, we extend the framework to the time-dependent case. We present an all-at-once discretization and develop a corresponding parallel-in-time preconditioner. In Section 4, we provide a spectral analysis of the preconditioners. In Section 5, we demonstrate the efficiency of the methods by a set of numerical experiments. Finally, in Section 6 we conclude and summarize our work.

2. Steady-state problem.

2.1. Problem formulation. Following the standard framework [11, 29], let $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space with $\xi : \Omega \rightarrow \Phi \subset \mathbb{R}^{m_\xi}$ a vector of independent random variables. We consider the Hilbert space $L^2(\Phi)$ equipped with the inner product $\langle u, v \rangle = \mathbb{E}[uv]$, where \mathbb{E} denotes the mathematical expectation with respect to the probability measure μ induced by ξ .

We consider the steady-state optimal control problem given by

$$(2.1) \quad \min_{y, u} J(y, u) := \frac{1}{2} \int_{\Phi} \int_{\mathcal{D}} |y - y_d|^2 dx d\mu(\xi) + \frac{\beta}{2} \int_{\Phi} \int_{\mathcal{D}} |u|^2 dx d\mu(\xi) + \frac{\gamma}{2} \int_{\mathcal{D}} |\sigma(y)|^2 dx,$$

subject to

$$(2.2) \quad \begin{cases} -\nabla \cdot (\mathbb{k}(x, \xi) \nabla y(x, \xi)) = u(x, \xi), & \text{in } \mathcal{D} \times \Phi, \\ y(x, \xi) = g(x), & \text{on } \partial\mathcal{D} \times \Phi, \end{cases}$$

where y is the state, y_d is the target state, and u is the (distributed) control. The parameter γ penalizes the variance $\sigma^2(y)$ of the state y . Observe that both the state y and control u are stochastic. In view of the Doob-Dynkin lemma (see, e.g., [2]), both y and u admit the same parametric dependence on ξ .

In computations, we work with a finite dimensional subspace $\mathcal{T}_p \subset L^2(\Phi, \mathcal{B}(\Phi), \mu)$, spanned by a set of generalized polynomial chaos (gPC) functions $\{\psi_\ell(\xi)\}_{\ell=1}^{n_A}$ with

$$(2.3) \quad \langle \psi_\ell, \psi_k \rangle = \mathbb{E}[\psi_\ell \psi_k] = \int_{\Phi} \psi_\ell(\xi) \psi_k(\xi) d\mu = \delta_{\ell k},$$

where $\{\psi_i(\xi)\}_{i=1}^{n_A}$ is a set of m_ξ -dimensional, p -order Hermite polynomials. $\psi_1(\xi) = 1$, and $\mathbb{E}[\psi_k(\xi)] = 0$ for $k > 1$.

It is assumed that $\mathbb{k}(x, \xi)$ is bounded away from zero and infinity, i.e., $0 < \mathbb{k}_{\min} \leq \mathbb{k}(x, \xi) \leq \mathbb{k}_{\max} < \infty$ for some constants $\mathbb{k}_{\min}, \mathbb{k}_{\max}$. We assume the stochastic PDE coefficient $\mathbb{k}(x, \xi)$ is represented by a gPC expansion

$$(2.4) \quad \mathbb{k}(x, \xi) = \sum_{i=1}^{n_A} \kappa_i(x) \psi_i(\xi),$$

Similarly, applying the stochastic Galerkin method to (2.1)–(2.2), both the state y and control u are expanded as

$$(2.5) \quad v = \sum_{i=1}^{n_h} \sum_{k=1}^{n_\xi} v_{ik} \phi_i(x) \psi_k(\xi), \quad v = y \text{ or } u.$$

Let $f(x, \xi) = f_0(x) + \sum_{i=1}^{n_\xi} \sqrt{\theta_i} f_i(x) \xi_i$ be a truncated Karhunen-Loève(KL) expansion of a Gaussian process defined on \mathcal{D} , where ξ_i are independent, identically distributed random variables, $f_0(x)$ is the mean function, and $(\theta_i, f_i(x))$ is the i -th eigenpair (eigenvalue and eigenfunctions, respectively) of the covariance function $C_f(x, y)$ such that

$$(2.6) \quad \int_{\mathcal{D}} C_f(x, y) \varphi_i(y) dy = \theta_i \varphi_i(x),$$

where $f_i(x) = \sqrt{\theta_i} \varphi_i(x)$. In particular, if $\mathbb{k}(x, \xi) = \exp[f(x, \xi)]$, then following [10]

$$\kappa_i(x) = \frac{\mathbb{E}[\psi_i(\xi - f)]}{\mathbb{E}[\psi_i^2]} \exp \left[f_0(x) + \frac{1}{2} \sum_{j=1}^{m_\xi} f_j^2(x) \right].$$

According to [18], to guarantee a complete representation of the lognormal random field, the variables y and u are defined in the space \mathcal{T}_p with dimension $n_\xi = \binom{m_\xi + p}{p}$, and the dimension of $\mathbb{k}(x, \xi)$ is given by $n_A = \binom{m_\xi + 2p}{2p}$.

Using the stochastic Galerkin finite element framework to discretize problem (2.1)–(2.2), we get

$$\min_{\mathbf{y}, \mathbf{u}} J(\mathbf{y}, \mathbf{u}) = \frac{1}{2} (\mathbf{y} - \mathbf{y}_d)^T \mathcal{M} (\mathbf{y} - \mathbf{y}_d) + \frac{\gamma}{2} \mathbf{y}^T \mathcal{M}_\sigma \mathbf{y} + \frac{\beta}{2} \mathbf{u}^T \mathcal{M} \mathbf{u},$$

subject to

$$(2.7) \quad -\mathcal{A} \mathbf{y} + \mathcal{M} \mathbf{u} = \mathbf{g}.$$

Via the Lagrangian formulation, we can derive the first-order optimality conditions

$$(2.8) \quad \begin{bmatrix} \mathcal{M}_\gamma & 0 & -\mathcal{A}^T \\ 0 & \beta \mathcal{M} & \mathcal{M}^T \\ -\mathcal{A} & \mathcal{M} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathcal{M} \mathbf{y}_d \\ 0 \\ \mathbf{g} \end{bmatrix},$$

where $\boldsymbol{\lambda}$ is a Lagrangian multiplier, $\mathcal{M}_\gamma = \mathcal{M} + \gamma \mathcal{M}_\sigma = H^\gamma \otimes M$, $H^\gamma = H_1 + \gamma H^\sigma$, and similarly, h^γ is defined as the entry of H^γ . In particular,

$$(2.9) \quad \mathcal{M} = H_1 \otimes M, \quad \mathcal{M}_\sigma = H^\sigma \otimes M,$$

where the mass matrix, M is given by

$$M_{a,b} = \int_D \phi_a(x)\phi_b(x)dx, \quad a, b = 1, 2, \dots, n_h,$$

for a suitably chosen finite element basis, $\{\phi_i(x)\}_{i=1}^{n_h}$. The global stiffness matrix \mathcal{A} is given by $\mathcal{A} = \sum_{\ell=1}^{n_A} H_\ell \otimes A_\ell$. Here, the matrices H_ℓ are defined by their entries $(H_\ell)_{jk} = h_{\ell jk}$, with $h_{ijk} = \mathbb{E}[\psi_i\psi_j\psi_k]$, where we note that all matrices H_ℓ are symmetric, and

$$H^\sigma = \text{diag}(0, h_{1,jj}), \quad j = 2, \dots, n_\xi,$$

which is obtained from H_1 by setting $h_{1,11} = 0$. Note also that in our settings $H_1 = I_{n_\xi}$, i.e., it is an identity matrix of size n_ξ . Moreover, all matrices A_ℓ share the same sparsity pattern and are given by

$$(2.10) \quad A_\ell = [(A_\ell)_{ab}], \quad (A_\ell)_{ab} = \int_D \kappa_\ell(x) \nabla \phi_a(x) \cdot \nabla \phi_b(x) dx, \quad a, b = 1, 2, \dots, n_h.$$

We note that the coefficient matrix in (2.8) is symmetric, indefinite, and in general very large. It is ill-conditioned, and therefore, a good preconditioner is required to solve the system efficiently. Next, we introduce a preconditioner to tackle this problem.

2.2. Schur complement-based preconditioner. A block-diagonal preconditioner for (2.8) (see, e.g. Benner et al. [5]) is given by

$$(2.11) \quad \mathcal{P} := \begin{bmatrix} \mathcal{M}_\gamma & 0 & 0 \\ 0 & \beta \mathcal{M} & 0 \\ 0 & 0 & \mathcal{S}_{\text{exact}} \end{bmatrix},$$

where $\mathcal{S}_{\text{exact}}$ is the exact Schur complement

$$(2.12) \quad \mathcal{S}_{\text{exact}} = \mathcal{A} \mathcal{M}_\gamma^{-1} \mathcal{A}^T + \frac{1}{\beta} \mathcal{M}.$$

The first two blocks corresponding to (scaling of) the mass matrix are block diagonal, and the inverses are approximated by Chebyshev semi-iteration. However, forming and applying the inverse of $\mathcal{S}_{\text{exact}}$ is computationally prohibitive. The primary difficulty stems from its additive structure. Therefore, the key to an efficient solution lies in designing an approximation \mathcal{S} that is spectrally equivalent to $\mathcal{S}_{\text{exact}}$ and easy to invert. Following the approach in [5], we employ the approximation

$$(2.13) \quad \mathcal{S} = \mathcal{Z} \mathcal{M}_\gamma^{-1} \mathcal{Z}^T, \quad \mathcal{Z} = \mathcal{A} + \sqrt{\frac{1+\gamma}{\beta}} \mathcal{M} = \sum_{\ell=1}^{n_A} H_\ell \otimes \tilde{A}_\ell,$$

where

$$\tilde{A}_1 = A_1 + \sqrt{\frac{1+\gamma}{\beta}} M, \quad \tilde{A}_\ell = A_\ell, \quad \ell = 2, \dots, n_A.$$

Since \mathcal{Z} is symmetric, we have $\mathcal{S}^{-1} = \mathcal{Z}^{-1} \mathcal{M}_\gamma \mathcal{Z}^{-1}$. In [5], the authors studied a mean-based preconditioner derived from (2.13) by dropping all the terms of \mathcal{Z} except the first term; that is, $\mathcal{Z} \approx H_1 \otimes \tilde{A}_1$. However, this mean-based preconditioner is less effective when the standard deviation of the uncertain parameters increases.

In this study, we overcome this shortcoming by hierarchical preconditioning introduced in [25, 26] in the context of forward problems. More specifically, unlike

[25, 26], we extend this strategy to the stochastic optimal control problem, and we also study the spectral properties of the new preconditioner.

To that end, observe first that, by applying the identity

$$(2.14) \quad (V \otimes W) \text{vec}(X) = \text{vec}(W X V^T),$$

to (2.8) yields the matricized system [5, eq. (53)]:

$$(2.15) \quad \begin{aligned} M \bar{Y} H^\gamma - \sum_{\ell=1}^{n_A} \tilde{A}_\ell^T \bar{\Lambda} H_\ell &= M \bar{Y}_d H_1, \\ \beta M \bar{U} H_1 + M^T \bar{\Lambda} H_1 &= 0, \\ - \sum_{\ell=1}^{n_A} \tilde{A}_\ell^T \bar{Y} H_\ell + M \bar{U} H_1 &= \bar{G}. \end{aligned}$$

In addition, the matrix-vector of $\mathcal{Z}\bar{v}$ in (2.13) reads

$$(2.16) \quad \mathcal{Z}\bar{v} = \text{vec} \left(\sum_{\ell=1}^{n_A} \tilde{A}_\ell \bar{V} H_\ell \right).$$

2.3. Hierarchical Gauss-Seidel preconditioner. We first recall the preconditioner for the forward PDE problem from [25]. However, in this work, we present it in the matricized format as Algorithm 2.1. To set the notation, we note that the *matricized* format, which utilizes the isomorphism between $\mathbb{R}^{n_h n_\xi}$ and $\mathbb{R}^{n_h \times n_\xi}$, defined via the operators vec and mat . Specifically,

$$(2.17) \quad \bar{V} = \text{mat}(\bar{v}) = [v_1, v_2, \dots, v_{n_\xi}] \in \mathbb{R}^{n_h \times n_\xi},$$

where the column k contains the coefficients associated with the basis function ψ_k , and $\bar{v} = \text{vec}(\bar{V}) \in \mathbb{R}^{n_h n_\xi}$. We will use lowercase letters for the *vectorized* representation and uppercase letters for the *matricized* counterpart; so, e.g., $\bar{R} = \text{mat}(\bar{r})$, etc.

Moreover, we will denote by $\bar{V}_{(i:n)}$ a submatrix of \bar{V} containing columns $i, i+1, \dots, n$, and, in particular, $\bar{V} = \bar{V}_{(1:n_\xi)}$. There are two components of the preconditioner. The first component consists of block-diagonal solves with blocks of varying sizes. The second component is used in the setup of the right-hand sides for the solves, and consists of matrix-vector products by certain subblocks of the stochastic Galerkin matrix by vectors of corresponding sizes. To this end, we will write $[h_{\tau,(\ell)(k)}]$, with (ℓ) and (k) denoting a set of (consecutive) rows and columns of matrix H_τ so that, in particular, $H_\tau = [h_{\tau,(1:n_\xi)(1:n_\xi)}]$. Let us also denote $\bar{v}_{(\ell)} = \text{vec}(\bar{V}_{(\ell)})$. Then, the matrix-vector products can be written, cf. (2.14) and noting the symmetry of H_τ , as

$$(2.18) \quad \bar{v}_{(\ell)} = \sum_{\tau \in \mathcal{I}_\tau} ([h_{\tau,(\ell)(k)}] \otimes \tilde{A}_\tau) \bar{u}_{(k)} \Leftrightarrow \bar{V}_{(\ell)} = \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{U}_{(k)} [h_{\tau,(k)(\ell)}],$$

where \mathcal{I}_τ is an index set $\mathcal{I}_\tau \subseteq \{1, \dots, n_\tau\}$ indicating that the matrix-vector products may be truncated. Possible strategies for truncation are discussed in [25]. In this study, we use $\mathcal{I}_\tau = \{1, \dots, n_\tau\}$ with $n_\tau = \binom{m_\xi + p_\tau}{p_\tau}$ for some $p_\tau \leq p$. In particular, we set $\tau = \{0, 1, 2\}$ and with $\mathcal{I}_\tau = \emptyset$ both preconditioners in Algorithms 2.1 and 2.2–2.3 reduce to mean-based variants. We also note that, since the initial guess is zero in Algorithm 2.1, the multiplications by \mathcal{F}_1 and \mathcal{F}_{d+1} vanish from (2.19)–(2.20).

Next, we apply the hGS strategy to form a preconditioner for the KKT system. The preconditioner is formulated as Algorithm 2.2–2.3. It adapts the core idea of Algorithm 2.1 to handle the coupled variables corresponding to the state \bar{V}^Y , control \bar{V}^U ,

Algorithm 2.1 [25, Algorithm 3] Hierarchical Gauss-Seidel preconditioner (hGS)

The preconditioner $\mathcal{Z}_{hGS} : \bar{R} \mapsto \bar{V}$ is defined as follows.

- 1: Set the initial solution \bar{V} to zero and update in the following steps:
- 2: Solve

$$(2.19) \quad \tilde{A}_1 \bar{V}_{(1)} = \bar{R}_{(1)} - \mathcal{F}_1, \quad \text{where } \mathcal{F}_1 = \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{V}_{(2:n_\xi)} [h_{\tau, (2:n_\xi)(1)}].$$

- 3: **for** $d = 1, \dots, p-1$ **do**

- 4: Set $\ell = (n_\ell + 1 : n_u)$, where $n_\ell = \binom{m_\xi + d - 1}{d-1}$ and $n_u = \binom{m_\xi + d}{d}$.

- 5: Solve

$$(2.20) \quad \tilde{A}_1 \bar{V}_{(\ell)} = \bar{R}_{(\ell)} - \mathcal{E}_{d+1} - \mathcal{F}_{d+1},$$

where

$$\mathcal{E}_{d+1} = \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{V}_{(1:n_\ell)} [h_{\tau, (1:n_\ell)(\ell)}], \quad \mathcal{F}_{d+1} = \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{V}_{(n_u+1:n_\xi)} [h_{\tau, (n_u+1:n_\xi)(\ell)}].$$

- 6: **end for**

- 7: Set $\ell = (n_u + 1 : n_\xi)$.

- 8: Solve

$$\tilde{A}_1 \bar{V}_{(\ell)} = \bar{R}_{(\ell)} - \mathcal{E}_{p+1}, \quad \text{where } \mathcal{E}_{p+1} = \sum_{t \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{V}_{(1:n_u)} [h_{\tau, (1:n_u)(\ell)}],$$

- 9: **for** $d = p-1, \dots, 1$ **do**

- 10: Set $\ell = (n_\ell + 1 : n_u)$, where $n_\ell = \binom{m_\xi + d - 1}{d-1}$ and $n_u = \binom{m_\xi + d}{d}$.

- 11: Solve (2.20).

- 12: **end for**

- 13: Solve (2.19).

and adjoint \bar{V}^Λ simultaneously at each hierarchical level. A key computational step of Algorithm 2.2–2.3 entails a solve with $\tilde{P} = \text{blkdiag} [M, \beta M, \tilde{A}_1]$, which serves as an auxiliary block-diagonal preconditioner derived from the blocks of the deterministic KKT system. Since this system is relatively small and constant across all hierarchical levels, it can be handled efficiently, for instance, by computing a direct factorization of \tilde{P} once and reusing it for all subsequent solves. The overall performance of the preconditioner is thus determined by the cost of these deterministic solves and the number of truncated off-diagonal matrix-vector products.

3. Time-dependent problem. The time-dependent optimal control problem is given by

$$(3.1) \quad \begin{aligned} \min_{y,u} \mathcal{J}(y,u) &= \frac{1}{2} \int_0^T \int_{\Phi} \int_{\mathcal{D}} |y - y_d|^2 dx d\mu(\xi) dt + \frac{\beta}{2} \int_0^T \int_{\Phi} \int_{\mathcal{D}} |u|^2 dx d\mu(\xi) dt \\ &+ \frac{\gamma}{2} \int_0^T \int_{\Phi} \int_{\mathcal{D}} |\sigma(y)|^2 dx d\mu(\xi) dt, \end{aligned}$$

Algorithm 2.2 hGS preconditioner for the optimal control problem (hGSoc)

The preconditioner $\mathcal{P}_{hGSoc} : (\bar{R}^Y, \bar{R}^U, \bar{R}^\Lambda) \mapsto (\bar{V}^Y, \bar{V}^U, \bar{V}^\Lambda)$ is defined as follows.

- 1: Set the initial solution $(\bar{V}^Y, \bar{V}^U, \bar{V}^\Lambda)$ to zero and update in the following steps:
- 2: Solve

$$(2.21) \quad \tilde{P} \begin{bmatrix} \bar{V}_{(1)}^Y \\ \bar{V}_{(1)}^U \\ \bar{V}_{(1)}^\Lambda \end{bmatrix} = \begin{bmatrix} \bar{R}_{(1)}^Y - \mathcal{C}_1 + \mathcal{D}_1 \\ \bar{R}_{(1)}^U - \mathcal{E}_1 - \mathcal{F}_1 \\ \bar{R}_{(1)}^\Lambda + \mathcal{G}_1 - \mathcal{H}_1 \end{bmatrix},$$

where

$$\begin{aligned} \mathcal{C}_1 &= M \bar{V}_{(2:n_\xi)}^Y \left[h_{(2:n_\xi)(1)}^\gamma \right], & \mathcal{D}_1 &= \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{V}_{(2:n_\xi)}^\Lambda \left[h_{\tau, (2:n_\xi)(1)} \right], \\ \mathcal{E}_1 &= \beta M \bar{V}_{(2:n_\xi)}^U \left[h_{1, (2:n_\xi)(1)} \right], & \mathcal{F}_1 &= M \bar{V}_{(2:n_\xi)}^\Lambda \left[h_{1, (2:n_\xi)(1)} \right], \\ \mathcal{G}_1 &= \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \bar{V}_{(2:n_\xi)}^Y \left[h_{\tau, (2:n_\xi)(1)} \right], & \mathcal{H}_1 &= M \bar{V}_{(2:n_\xi)}^U \left[h_{1, (2:n_\xi)(1)} \right]. \end{aligned}$$

- 3: **for** $d = 1, \dots, p-1$ **do**
- 4: Set $\ell = (n_\ell + 1 : n_u)$, where $n_\ell = \binom{m_\xi + d - 1}{d-1}$ and $n_u = \binom{m_\xi + d}{d}$.
- 5: Solve

$$(2.22) \quad \tilde{P} \begin{bmatrix} \bar{V}_{(\ell)}^Y \\ \bar{V}_{(\ell)}^U \\ \bar{V}_{(\ell)}^\Lambda \end{bmatrix} = \begin{bmatrix} \bar{R}_{(\ell)}^Y - \mathcal{C}_{d+1} + \mathcal{D}_{d+1} \\ \bar{R}_{(\ell)}^U - \mathcal{E}_{d+1} - \mathcal{F}_{d+1} \\ \bar{R}_{(\ell)}^\Lambda + \mathcal{G}_{d+1} - \mathcal{H}_{d+1} \end{bmatrix},$$

where

$$\begin{aligned} \mathcal{C}_{d+1} &= M \left(\bar{V}_{(1:n_\ell)}^Y \left[h_{(1:n_\ell)(\ell)}^\gamma \right] + \bar{V}_{(n_u+1:n_\xi)}^Y \left[h_{(n_u+1:n_\xi)(\ell)}^\gamma \right] \right), \\ \mathcal{D}_{d+1} &= \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \left(\bar{V}_{(1:n_\ell)}^\Lambda \left[h_{\tau, (1:n_\ell)(\ell)} \right] + \bar{V}_{(n_u+1:n_\xi)}^\Lambda \left[h_{\tau, (n_u+1:n_\xi)(\ell)} \right] \right), \\ \mathcal{E}_{d+1} &= \beta M \left(\bar{V}_{(1:n_\ell)}^U \left[h_{1, (1:n_\ell)(\ell)} \right] + \bar{V}_{(n_u+1:n_\xi)}^U \left[h_{1, (n_u+1:n_\xi)(\ell)} \right] \right), \\ \mathcal{F}_{d+1} &= M \left(\bar{V}_{(1:n_\ell)}^\Lambda \left[h_{1, (1:n_\ell)(\ell)} \right] + \bar{V}_{(n_u+1:n_\xi)}^\Lambda \left[h_{1, (n_u+1:n_\xi)(\ell)} \right] \right), \\ \mathcal{G}_{d+1} &= \sum_{\tau \in \mathcal{I}_\tau} \tilde{A}_\tau \left(\bar{V}_{(1:n_\ell)}^Y \left[h_{\tau, (1:n_\ell)(\ell)} \right] + \bar{V}_{(n_u+1:n_\xi)}^Y \left[h_{\tau, (n_u+1:n_\xi)(\ell)} \right] \right), \\ \mathcal{H}_{d+1} &= M \left(\bar{V}_{(1:n_\ell)}^U \left[h_{1, (1:n_\ell)(\ell)} \right] + \bar{V}_{(n_u+1:n_\xi)}^U \left[h_{1, (n_u+1:n_\xi)(\ell)} \right] \right). \end{aligned}$$

6: **end for**

subject to

$$(3.2) \quad \begin{cases} \frac{\partial y(t, \mathbf{x}, \xi)}{\partial t} - \nabla \cdot (\mathbb{k}(\mathbf{x}, \xi) \nabla y(t, \mathbf{x}, \xi)) = u(t, \mathbf{x}, \xi) \text{ in } (0, T] \times \mathcal{D} \times \Phi, \\ y(t, \mathbf{x}, \xi) = g \text{ on } (0, T] \times \partial \mathcal{D} \times \Phi, \\ y(0, \mathbf{x}, \xi) = y_0 \text{ in } \mathcal{D} \times \Phi. \end{cases}$$

Algorithm 2.3 hGS preconditioner for the optimal control problem (hGSoc), cont'd

7: Set $\ell = (n_u + 1 : n_\xi)$.

8: Solve

$$\tilde{P} \begin{bmatrix} \bar{V}_{(\ell)}^Y \\ \bar{V}_{(\ell)}^U \\ \bar{V}_{(\ell)}^\Lambda \end{bmatrix} = \begin{bmatrix} \bar{R}_{(\ell)}^Y - \mathcal{C}_{p+1} + \mathcal{D}_{p+1} \\ \bar{R}_{(\ell)}^U - \mathcal{E}_{p+1} - \mathcal{F}_{p+1} \\ \bar{R}_{(\ell)}^\Lambda + \mathcal{G}_{p+1} - \mathcal{H}_{p+1} \end{bmatrix},$$

where

$$\begin{aligned} \mathcal{C}_{p+1} &= M\bar{V}_{(1:n_u)}^Y [h_{(1:n_u)}^\gamma], & \mathcal{D}_{p+1} &= \sum_{\tau \in \mathcal{I}_\tau} A_\tau \bar{V}_{(1:n_u)}^\Lambda [h_{\tau, (1:n_u)}(\ell)], \\ \mathcal{E}_{p+1} &= \beta M\bar{V}_{(1:n_u)}^U [h_{1, (1:n_u)}(\ell)], & \mathcal{F}_{p+1} &= M\bar{V}_{(1:n_u)}^\Lambda [h_{1, (1:n_u)}(\ell)], \\ \mathcal{G}_{p+1} &= \sum_{\tau \in \mathcal{I}_\tau} A_\tau \bar{V}_{(1:n_u)}^Y [h_{\tau, (1:n_u)}(\ell)], & \mathcal{H}_{p+1} &= M\bar{V}_{(1:n_u)}^U [h_{1, (1:n_u)}(\ell)]. \end{aligned}$$

9: **for** $d = p - 1, \dots, 1$ **do**

10: Set $\ell = (n_\ell + 1 : n_u)$, where $n_\ell = \binom{m_\xi + d - 1}{d - 1}$ and $n_u = \binom{m_\xi + d}{d}$.

11: Solve (2.22).

12: **end for**

13: Solve (2.21).

After the application of the stochastic Galerkin finite element discretization to (3.1), and using the trapezoidal rule for the time discretization, where $n_t = T/\tau$ is the number of time steps over the interval $[0, T]$ with time-step size τ , we obtain

$$(3.3) \quad \min_{\mathbf{y}, \mathbf{u}} \mathcal{J}(\mathbf{y}, \mathbf{u}) = \frac{\tau}{2} (\mathbf{y} - \mathbf{y}_d)^T (D \otimes \mathcal{M}_\gamma) (\mathbf{y} - \mathbf{y}_d) + \frac{\tau\beta}{2} \mathbf{u}^T (D \otimes \mathcal{M}) \mathbf{u},$$

where \mathcal{M} and \mathcal{M}_γ are defined in (2.9), D in (3.6) below, \mathbf{y} , \mathbf{y}_d , and \mathbf{u} are vectors corresponding to the state, desired state, and control, respectively, that contain concatenated vectors $\mathbf{y}_i, \mathbf{y}_{d_i}, \mathbf{u}_i \in \mathbb{R}^{n_h n_\xi \times 1}, i = 1, \dots, n_t$, due to the time-stepping,

$$\mathbf{y} = [\mathbf{y}_1 \quad \dots \quad \mathbf{y}_{n_t}]^T, \mathbf{y}_d = [\mathbf{y}_{d_1} \quad \dots \quad \mathbf{y}_{d_{n_t}}]^T, \text{ and } \mathbf{u} = [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_{n_t}]^T.$$

After the application of the stochastic Galerkin finite element discretization to (3.2), and using the implicit Euler method for the time discretization, we obtain

$$(3.4) \quad M\mathbf{y}_k + \tau A\mathbf{y}_k = M\mathbf{y}_{k-1} + \tau M\mathbf{u}_k.$$

Combining all time steps of (3.4) in all-at-once discretization ([19, 21]), we can write

$$\mathcal{A}_t \mathbf{y} - \tau \mathcal{N} \mathbf{u} = [\mathcal{M} \mathbf{y}_0 + \mathbf{g}, \quad \mathbf{g}, \quad \dots, \mathbf{g}]^T =: \mathbf{d},$$

the vector \mathbf{g} contains the contributions from the Dirichlet boundary data at each time step, and

$$(3.5) \quad \mathcal{A}_t = (I_{n_t} \otimes \mathcal{L}) - (C \otimes \mathcal{M}), \mathcal{N} = I_{n_t} \otimes H_1 \otimes M, \text{ and } \mathcal{L} = H_1 \otimes (M + \tau A_1) + \tau \sum_{\ell=2}^{n_A} H_\ell \otimes A_\ell.$$

where the matrix C and matrix D , used in (3.3) and also below, are defined as

$$(3.6) \quad C = \text{subdiag}(1, 1, \dots, 1)_{-1}, \quad D = \text{diag}\left(\frac{1}{2}, 1, \dots, 1, \frac{1}{2}\right).$$

Forming the Lagrangean and applying the first-order optimality conditions, we get

$$(3.7) \quad \begin{bmatrix} \tau D \otimes \mathcal{M}_\gamma & 0 & -\mathcal{A}_t^T \\ 0 & \beta \tau D \otimes \mathcal{M} & \tau \mathcal{N}^T \\ -\mathcal{A}_t & \tau \mathcal{N} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \tau (D \otimes \mathcal{M}) \cdot (\mathbf{1}_{n_t} \otimes \mathbf{y}_d) \\ \mathbf{0} \\ \mathbf{d} \end{bmatrix},$$

where the boundary-inclusive source vector \mathbf{d} is defined as above, and $\mathbf{1}_{n_t} \in \mathbb{R}^{n_t \times 1}$ is the column all-ones vector.

3.1. PINT-based block-diagonal hierarchical Gauss-Seidel preconditioner. In analogy to (2.11), we propose a preconditioner for (3.7) as

$$(3.8) \quad \mathcal{P}_{\text{hGSoc-PINT}} \approx \begin{bmatrix} \tau D \otimes \mathcal{M}_\gamma & & \\ & \tau \beta D \otimes \mathcal{M} & \\ & & \bar{\mathcal{S}} \end{bmatrix},$$

where $\bar{\mathcal{S}}$ is a computationally efficient approximation of the exact Schur complement

$$(3.9) \quad \bar{\mathcal{S}}_{\text{exact}} = \frac{1}{\tau} \mathcal{A}_t (D \otimes \mathcal{M}_\gamma)^{-1} \mathcal{A}_t^T + \frac{\beta}{\tau} \mathcal{N} (D \otimes \mathcal{M})^{-1} \mathcal{N}^T.$$

The first two blocks that correspond to (scaling of) the mass matrix are approximated by Chebyshev semi-iteration. Since the iteration entails matrix-vector multiplications, we note that using (2.14) we have

$$\begin{aligned} (\tau D \otimes H_1 \otimes M) \mathbf{v}_1 &= \tau \left[\frac{1}{2} \text{vec}(MV_{11}H_\gamma) \quad \text{vec}(MV_{12}H_\gamma) \quad \cdots \quad \frac{1}{2} \text{vec}(MV_{1n_t}H_\gamma) \right]^T \\ (\beta \tau D \otimes H_1 \otimes M) \mathbf{v}_2 &= \beta \tau \left[\frac{1}{2} \text{vec}(MV_{21}H_1) \quad \text{vec}(MV_{22}H_1) \quad \cdots \quad \frac{1}{2} \text{vec}(MV_{2n_t}H_1) \right]^T, \end{aligned}$$

where \mathbf{v}_1 and \mathbf{v}_2 are the vectors obtained by concatenating \mathbf{v}_{1i} and \mathbf{v}_{2i} , $i = 1, \dots, n_t$, respectively, which correspond to the time steps. The matrices $V_{1i}, V_{2i} \in \mathbb{R}^{n_h \times n_\varepsilon}$ are then the *matricized* counterparts of \mathbf{v}_{1i} and \mathbf{v}_{2i} , respectively. Next, since inverting $\bar{\mathcal{S}}_{\text{exact}}$ is computationally prohibitive, we propose an approximation as

$$(3.10) \quad \bar{\mathcal{S}} = \frac{1}{\tau} \underbrace{\left(\mathcal{A}_t + \tau \sqrt{\frac{1+\gamma}{\beta}} \mathcal{N} \right)}_{=: \bar{\mathcal{Z}}} (D \otimes \mathcal{M}_\gamma)^{-1} \left(\mathcal{A}_t + \tau \sqrt{\frac{1+\gamma}{\beta}} \mathcal{N} \right)^T,$$

where using (3.5), we can rewrite $\bar{\mathcal{Z}}$ as

$$(3.11) \quad \bar{\mathcal{Z}} = I_{n_t} \otimes \left(H_1 \otimes (M + \tau A_1) + \tau \sum_{\ell=2}^{n_A} H_\ell \otimes A_\ell \right) + \tau \sqrt{\frac{1+\gamma}{\beta}} I_{n_t} \otimes H_1 \otimes M - C \otimes H_1 \otimes M.$$

By dropping the last term $C \otimes H_1 \otimes M$, we further approximate $\bar{\mathcal{Z}}$ by

$$(3.12) \quad \tilde{\mathcal{Z}} = I_{n_t} \otimes \left\{ H_1 \otimes \left[(1 + \tau \sqrt{\frac{1+\gamma}{\beta}}) M + \tau A_1 \right] + \tau \sum_{\ell=2}^{n_A} H_\ell \otimes A_\ell \right\}.$$

Observe that $\tilde{\mathcal{Z}}$ is symmetric, and it can be rewritten as

$$(3.13) \quad \tilde{\mathcal{Z}} = \tau \sqrt{\frac{1+\gamma}{\beta}} \mathcal{N} + I_{n_t} \otimes \mathcal{L}.$$

The idea is to use

$$(3.14) \quad \tilde{\mathcal{S}} \approx \tilde{\mathcal{Z}} (D \otimes \mathcal{M}_\gamma)^{-1} \tilde{\mathcal{Z}},$$

and in particular the solves with $\tilde{\mathcal{Z}}$ are approximated by Algorithm 2.1, similarly to the steady-state case. We remark that by dropping all terms with $\ell > 1$ from $\tilde{\mathcal{Z}}$ in (3.12); that is, considering

$$\tilde{\mathcal{Z}}_0 = I_{n_t} \otimes H_1 \otimes \left[\left(1 + \tau \sqrt{\frac{1+\gamma}{\beta}}\right) M + \tau A_1 \right],$$

we recover the mean-based preconditioner [5]. Since the application of Algorithm 2.1 entails matrix-vector multiplications, using (2.14) we formulate $\tilde{\mathcal{Z}} \mathbf{v}_3$ as

$$(3.15) \quad \begin{aligned} \tilde{\mathcal{Z}} \mathbf{v}_3 &= \left(I_{n_t} \otimes \left\{ H_1 \otimes \left[\left(1 + \tau \sqrt{\frac{1+\gamma}{\beta}}\right) M + \tau A_1 \right] + \tau \sum_{\ell=2}^{n_A} H_\ell \otimes A_\ell \right\} \right) \mathbf{v}_3 \\ &= \left[\sum_{\ell=1}^{n_A} \text{vec} \left(\hat{A}_\ell V_{31} H_\ell \right) \quad \sum_{\ell=1}^{n_A} \text{vec} \left(\hat{A}_\ell V_{32} H_\ell \right) \quad \cdots \quad \sum_{\ell=1}^{n_A} \text{vec} \left(\hat{A}_\ell V_{3n_t} H_\ell \right) \right]^T, \end{aligned}$$

where

$$(3.16) \quad \hat{A}_\ell = \begin{cases} \left(1 + \tau \sqrt{\frac{1+\gamma}{\beta}}\right) M + \tau A_1, & \ell = 1, \\ \tau A_\ell, & \ell = 2, \dots, n_A, \end{cases}$$

and $V_{3i} \in \mathbb{R}^{n_h \times n_\varepsilon}$ is the *matricized* form of the i -th block of \mathbf{v}_3 , with $i = 1, \dots, n_t$.

The practical implementation of the preconditioner for the time-dependent system leverages the inherent structure of the *all-at-once* formulation. As defined in (3.12), the core operator of the Schur complement preconditioner, $\tilde{\mathcal{Z}}$, is block-diagonal with respect to the time steps. This structure extends to the entire KKT system, which then makes the preconditioning easily parallelizable. Specifically, an application of the preconditioner $\mathcal{P}_{\text{hGSoc-PINT}}$ entails an application of the steady-state optimal control preconditioner $\mathcal{P}_{\text{hGSoc}}$ from Algorithm 2.2–2.3 to all time steps simultaneously, and so it represents *parallel-in-time* (PINT) approach. It is summarized as Algorithm 3.1.

Algorithm 3.1 Parallel-in-time hGSoc preconditioner (hGSoc-PINT)

The preconditioner $\mathcal{P}_{\text{hGSoc-PINT}} : \bar{\mathbf{R}} \mapsto \bar{\mathbf{V}}$ is defined as:

- 1: **for** $k = 1, \dots, n_t$ **do**
 - 2: Extract $(\bar{R}_k^Y, \bar{R}_k^U, \bar{R}_k^\Lambda)$. (the subvector k of $\bar{\mathbf{R}}$)
 - 3: Calculate $(\bar{V}_k^Y, \bar{V}_k^U, \bar{V}_k^\Lambda) = \mathcal{P}_{\text{hGSoc}}(\bar{R}_k^Y, \bar{R}_k^U, \bar{R}_k^\Lambda)$ (apply Algorithm 2.2–2.3)
 - 4: **end for**
 - 5: Concatenate $\{(\bar{V}_k^Y, \bar{V}_k^U, \bar{V}_k^\Lambda)\}_{k=1}^{n_t}$ into $\bar{\mathbf{V}}$.
-

4. Spectral analysis of the preconditioners. Since the steady-state optimal control problem can be viewed as a special case of the time-dependent formulation (with

$n_t = 1$), we focus on analyzing the time-dependent setting; the steady-state results then follow as a direct consequence. The all-at-once discretization, presented in Section 3, couples all time steps simultaneously, yielding a significantly larger KKT system than its steady-state counterpart. Our goal is to prove that the proposed parallel-in-time preconditioner, based on the hGSoc-PINT in Algorithm 3.1, is spectrally equivalent to the ideal (but computationally more expensive) preconditioner.

DEFINITION 4.1 (Spectral Equivalence). *Two symmetric positive definite (SPD) matrices A and B are said to be spectrally equivalent, denoted $A \sim B$, if there exist positive constants $a \leq b$, such that*

$$a\mathbf{v}^T B\mathbf{v} \leq \mathbf{v}^T A\mathbf{v} \leq b\mathbf{v}^T B\mathbf{v}$$

holds for all non-zero vectors \mathbf{v} . Equivalently, all eigenvalues of the preconditioned matrix $B^{-1}A$ are contained within the fixed interval, which means $\lambda(B^{-1}A) \subset [a, b]$.

The proof proceeds by establishing spectral equivalences

$$(4.1) \quad \bar{\mathcal{S}}_{\text{exact}} \sim \bar{\mathcal{S}} \sim \tilde{\mathcal{S}} \sim \tilde{\mathcal{S}}_r \sim \tilde{\mathcal{S}}_{\text{hGS-PINT}}.$$

Here, $\bar{\mathcal{S}}_{\text{exact}}$ denotes the exact Schur complement (3.9). The matrix $\bar{\mathcal{S}}$, defined in (3.10), serves as an approximation of $\bar{\mathcal{S}}_{\text{exact}}$, while $\tilde{\mathcal{S}}$ given in (3.14) is a block-diagonal approximation of $\bar{\mathcal{S}}$. The operator $\tilde{\mathcal{S}}_r$ represents the truncated hierarchical operator

$$(4.2) \quad \tilde{\mathcal{S}}_r = \tilde{\mathcal{Z}}_r (D \otimes \mathcal{M}_\gamma)^{-1} \tilde{\mathcal{Z}}_r^T, \quad \tilde{\mathcal{Z}}_r = I_{n_t} \otimes \left\{ H_1 \otimes \left[(1 + \tau \sqrt{\frac{1+\gamma}{\beta}})M + \tau A_1 \right] + \tau \sum_{\ell=2}^r H_\ell \otimes A_\ell \right\},$$

with $r = 1, \dots, n_A$. When $r = 1$, $\tilde{\mathcal{S}}_r$ reduces to the mean-based preconditioner employed in [5], and when $r = n_A$, it recovers the full operator $\tilde{\mathcal{S}}$. Finally, $\tilde{\mathcal{S}}_{\text{hGS-PINT}} = \tilde{\mathcal{Z}}_{\text{hGS-PINT}} (D \otimes \mathcal{M}_\gamma)^{-1} \tilde{\mathcal{Z}}_{\text{hGS-PINT}}^T$ represents the computationally feasible approximation of $\tilde{\mathcal{S}}_r$ in which the linear systems $\tilde{\mathcal{Z}}_r \mathbf{x} = \mathbf{b}$ are solved approximately via the hierarchical Gauss-Seidel method (Algorithm 2.1), as implemented in the parallel-in-time framework of Algorithm 3.1.

To establish the spectral equivalences in this chain, we require several technical results. We state some auxiliary lemmas concerning matrix perturbations and congruence transformations, which will serve as building blocks for the main theorems. Their proofs are provided in Appendix A. In what follows, we denote by $\sigma_{\min}(\cdot)$ and $\sigma_{\max}(\cdot)$ the smallest and largest singular values, respectively.

LEMMA 4.2 (Spectral Equivalence via Error Bound). *Let A and B be symmetric positive definite matrices. Define the error matrix $E = A - B$. If there exists a constant $0 < \delta < 1$ such that for all non-zero vectors \mathbf{v} ,*

$$|\mathbf{v}^T E\mathbf{v}| \leq \delta(\mathbf{v}^T B\mathbf{v}),$$

then A and B are spectrally equivalent with

$$1 - \delta \leq \lambda(B^{-1}A) \leq 1 + \delta.$$

LEMMA 4.3 (Eigenvalues under Congruence Transformation). *Let C and D be symmetric positive definite matrices, and let Q be a nonsingular matrix. The eigenvalues of the pair (C, D) are identical to those of the transformed pair $(Q^T C Q, Q^T D Q)$.*

LEMMA 4.4. *For any nonzero vector \mathbf{v} , we have $\sigma_{\min}(A)\|\mathbf{v}\|_2 \leq \|A\mathbf{v}\|_2$.*

LEMMA 4.5. *For any matrices A and B of the same dimensions,*

$$\sigma_{\min}(A + B) \geq \sigma_{\min}(B) - \|A\|_2.$$

With these auxiliary results in place, we now proceed to establish the spectral equivalences in (4.1). We begin by recalling the following spectral equivalence result from [5].

THEOREM 4.6 ([5], Theorems 4, 6). *Let $\bar{\mathcal{S}}_{exact}$ be the exact Schur complement and $\bar{\mathcal{S}}$ be its approximation as given by (3.9) and (3.10), respectively. The eigenvalues of $\bar{\mathcal{S}}^{-1}\bar{\mathcal{S}}_{exact}$ are given by*

$$\lambda(\bar{\mathcal{S}}^{-1}\bar{\mathcal{S}}_{exact}) \subseteq \left[\frac{1}{2(1+\alpha)}, 1 \right),$$

where α satisfies $\alpha < \left(\frac{\sqrt{\kappa(\mathcal{A}_t)+1}}{\sqrt{\kappa(\mathcal{A}_t)-1}} \right)^2 - 1$, and \mathcal{A}_t is defined in (3.5).

Next, we show the relationship between $\bar{\mathcal{S}}$ and $\tilde{\mathcal{S}}$, as given by (3.10) and (3.14), respectively. First, however, we prove the following lemma.

LEMMA 4.7. *Let $W = (D \otimes \mathcal{M}_\gamma)^{-1}$, and assume there exists a constant $\mu > 1$ such that*

$$(4.3) \quad \tau \sqrt{\frac{1+\gamma}{\beta}} \sigma_{\min}(\mathcal{N}W^{\frac{1}{2}}) \geq \mu \|(I_{n_t} \otimes \mathcal{L})W^{\frac{1}{2}}\|_2,$$

where \mathcal{L} is defined in (3.5). Then the minimum eigenvalue of $\tilde{\mathcal{S}}$ has the following lower bound

$$\lambda_{\min}(\tilde{\mathcal{S}}) \geq \frac{\tau}{\beta} \left(1 - \frac{1}{\mu}\right)^2 \sigma_{\min}^2(M^{\frac{1}{2}}).$$

Proof. Using the definition of $\tilde{\mathcal{S}}$ in (3.14), and the fact that for any real matrix A , $\lambda(AA^T) = \sigma_{\min}^2(A)$, we have

$$\lambda_{\min}(\tilde{\mathcal{S}}) = \lambda_{\min}(\tilde{\mathcal{Z}}W\tilde{\mathcal{Z}}^T) = \lambda_{\min}((\tilde{\mathcal{Z}}W^{\frac{1}{2}})(\tilde{\mathcal{Z}}W^{\frac{1}{2}})^T) = \sigma_{\min}^2(\tilde{\mathcal{Z}}W^{\frac{1}{2}}).$$

Recall from (3.13) that $\tilde{\mathcal{Z}}W^{\frac{1}{2}} = \tau \sqrt{\frac{1+\gamma}{\beta}} \mathcal{N}W^{\frac{1}{2}} + (I_{n_t} \otimes \mathcal{L})W^{\frac{1}{2}}$. Now, applying Lemma 4.5 yields

$$(4.4) \quad \sigma_{\min}(\tilde{\mathcal{Z}}W^{\frac{1}{2}}) \geq \tau \sqrt{\frac{1+\gamma}{\beta}} \sigma_{\min}(\mathcal{N}W^{\frac{1}{2}}) - \|(I_{n_t} \otimes \mathcal{L})W^{\frac{1}{2}}\|_2.$$

Now using (4.4) and the assumption (4.3), we obtain ¹

$$\sigma_{\min}(\tilde{\mathcal{Z}}W^{\frac{1}{2}}) \geq \left(1 - \frac{1}{\mu}\right) \tau \sqrt{\frac{1+\gamma}{\beta}} \sigma_{\min}(\mathcal{N}W^{\frac{1}{2}}).$$

Consequently, the lower bound for $\lambda_{\min}(\tilde{\mathcal{S}})$ is

$$\lambda_{\min}(\tilde{\mathcal{S}}) \geq \left(1 - \frac{1}{\mu}\right)^2 \left(\tau \sqrt{\frac{1+\gamma}{\beta}}\right)^2 \sigma_{\min}^2(\mathcal{N}W^{\frac{1}{2}}).$$

¹In our experience, this condition is often satisfied numerically when $\tau \gg \sqrt{\beta}$.

Next, observe that

$$\sigma_{\min}^2(\mathcal{N}W^{\frac{1}{2}}) = \sigma_{\min}^2\left((I_{n_t} \otimes \mathcal{M})(D^{-\frac{1}{2}} \otimes \mathcal{M}_\gamma^{-\frac{1}{2}})\right) = \sigma_{\min}^2(D^{-\frac{1}{2}}) \cdot \sigma_{\min}^2(\mathcal{M}\mathcal{M}_\gamma^{-\frac{1}{2}}),$$

where the last equality follows from the property that the singular values of a Kronecker product are the products of the singular values of the factors, i.e., $\sigma(A \otimes B) = \sigma(A)\sigma(B)$.

It is easy to verify that $\sigma_{\min}(D^{-\frac{1}{2}}) = 1$ and

$$\sigma_{\min}^2(\mathcal{M}\mathcal{M}_\gamma^{-\frac{1}{2}}) = \sigma_{\min}^2((H_1 \otimes M)(H_\gamma^{-\frac{1}{2}} \otimes M^{-\frac{1}{2}})) = \frac{1}{1+\gamma} \sigma_{\min}^2(M^{\frac{1}{2}}).$$

Thus, we have

$$\sigma_{\min}^2(\mathcal{N}W^{\frac{1}{2}}) = \frac{1}{1+\gamma} \sigma_{\min}^2(M^{\frac{1}{2}}).$$

Substituting this into the expression for the lower bound of $\lambda_{\min}(\tilde{\mathcal{S}})$, we finally get

$$\lambda_{\min}(\tilde{\mathcal{S}}) \geq \left(1 - \frac{1}{\mu}\right)^2 \frac{\tau^2(1+\gamma)}{\beta} \left(\frac{1}{1+\gamma} \sigma_{\min}^2(M^{\frac{1}{2}})\right) = \frac{\tau}{\beta} \left(1 - \frac{1}{\mu}\right)^2 \sigma_{\min}^2(M^{\frac{1}{2}}). \quad \square$$

THEOREM 4.8. *Assume that the conditions of Lemma 4.7 hold. Then, the eigenvalues $\lambda(\tilde{\mathcal{S}}^{-1}\tilde{\mathcal{S}})$ satisfy*

$$(1 - \theta)^2 \leq \lambda(\tilde{\mathcal{S}}^{-1}\tilde{\mathcal{S}}) \leq (1 + \theta)^2,$$

where θ is the perturbation parameter, with

$$\theta := \sup_{\mathbf{v} \neq 0} \frac{\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2}{\|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2} \leq \frac{\sqrt{2\beta}}{\sqrt{1+\gamma} \left(1 - \frac{1}{\mu}\right) \tau} \kappa(M^{\frac{1}{2}}).$$

Proof. Recall from (3.11) and (3.12), $\tilde{\mathcal{Z}} = \tilde{\mathcal{Z}} - C \otimes \mathcal{M}$, which implies for any $\mathbf{v} \neq 0$,

$$W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v} = W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v} - W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}.$$

Using the triangle inequality, we have

$$\left| \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2 - \|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2 \right| \leq \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2 \leq \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2 + \|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2.$$

From the definition of θ , we have $\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2 \leq \theta \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2$. Substituting this into the inequality above, we obtain

$$(1 - \theta) \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2 \leq \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2 \leq (1 + \theta) \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2.$$

Squaring these inequalities leads to the spectral bounds

$$(1 - \theta)^2 \leq \lambda(\tilde{\mathcal{S}}^{-1}\tilde{\mathcal{S}}) \leq (1 + \theta)^2.$$

Next, it remains to analyze the upper bound of θ . To this end, using Lemma 4.4, we know that $\sigma_{\min}(W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T) \|\mathbf{v}\|_2 \leq \|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2$.

$$\theta = \sup_{\mathbf{v} \neq 0} \frac{\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2}{\|W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T \mathbf{v}\|_2} \leq \frac{\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2}{\sigma_{\min}(W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T) \|\mathbf{v}\|_2} = \frac{\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T \mathbf{v}\|_2}{\sigma_{\min}(W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T)},$$

For the numerator, we have

$$\begin{aligned}\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T\|_2 &= \|(D^{-\frac{1}{2}} \otimes \mathcal{M}_\gamma^{-\frac{1}{2}})(C^T \otimes \mathcal{M}^T)\|_2 \\ &= \|(D^{-\frac{1}{2}}C^T) \otimes (\mathcal{M}_\gamma^{-\frac{1}{2}}\mathcal{M}^T)\|_2 \\ &= \|D^{-\frac{1}{2}}C^T\|_2 \cdot \|\mathcal{M}_\gamma^{-\frac{1}{2}}\mathcal{M}^T\|_2.\end{aligned}$$

Furthermore, the spectral norm of the Kronecker product satisfies

$$\|D^{-\frac{1}{2}}C^T\|_2 \cdot \|\mathcal{M}_\gamma^{-\frac{1}{2}}\mathcal{M}^T\|_2 \leq \frac{\sqrt{2}}{\sqrt{1+\gamma}}\|M^{\frac{1}{2}}\|_2,$$

which follows from the bound $\|D^{-1/2}C^T\|_2 \leq 1$ and the fact that $\|\mathcal{M}_\gamma^{-1/2}\mathcal{M}^T\|_2$ is bounded by the scaled norm of the square root of the mass matrix. Thus, the numerator satisfies

$$\|W^{\frac{1}{2}}(C \otimes \mathcal{M})^T\|_2 \leq \frac{\sqrt{2}}{\sqrt{1+\gamma}}\|M^{\frac{1}{2}}\|_2.$$

For the denominator, using lemma 4.7, we have

$$\sigma_{\min}(W^{\frac{1}{2}}\tilde{\mathcal{Z}}^T) = \sigma_{\min}(\tilde{\mathcal{Z}}W^{\frac{1}{2}}) \geq \left(1 - \frac{1}{\mu}\right) \frac{\tau}{\sqrt{\beta}}\sigma_{\min}(M^{\frac{1}{2}}).$$

Combining the bounds for the numerator and denominator,

$$\begin{aligned}\theta &\leq \frac{\frac{\sqrt{2}}{\sqrt{1+\gamma}}\|M^{\frac{1}{2}}\|_2}{\left(1 - \frac{1}{\mu}\right) \frac{\tau}{\sqrt{\beta}}\sigma_{\min}(M^{\frac{1}{2}})} \\ &= \frac{\sqrt{2\beta}}{\sqrt{1+\gamma}\left(1 - \frac{1}{\mu}\right)\tau} \cdot \frac{\sigma_{\max}(M^{\frac{1}{2}})}{\sigma_{\min}(M^{\frac{1}{2}})} \\ &= \frac{\sqrt{2\beta}}{\sqrt{1+\gamma}\left(1 - \frac{1}{\mu}\right)\tau} \kappa(M^{\frac{1}{2}}),\end{aligned}$$

thereby completing the proof of the theorem. \square

Note that the bound on θ shows that for sufficiently large τ (e.g., $\tau \gtrsim \sqrt{\beta} \kappa(M)^{1/2}$), we have $\theta < 1$; hence, $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}_r$ are spectrally equivalent. In practice, $\kappa(M)$ can be kept $\mathcal{O}(1)$ via appropriate basis choices and mass lumping, so it suffices to require $\tau \gg \sqrt{\beta}$.

Next, we establish the spectral equivalence between $\tilde{\mathcal{S}}$ and its truncated form $\tilde{\mathcal{S}}_r$. To this end, we shall rely on the error matrix $\mathcal{E}_r = \tilde{\mathcal{S}} - \tilde{\mathcal{S}}_r$, and the result of Lemma 4.2.

THEOREM 4.9 (Truncation Error Bound). *Let $\tilde{\mathcal{S}}$ be the full operator and $\tilde{\mathcal{S}}_r$ be the truncated hierarchical operator with truncation parameter r defined in (3.14) and (4.2), respectively. Assume $\nu := \lambda_{\min}(\hat{A}_1) > 0$. Define the truncation tail sum as*

$$\delta_r := \frac{1}{\nu} \sum_{\ell=r+1}^{n_A} \|H_\ell\|_2 \|\hat{A}_\ell\|_2.$$

If the decay of the stochastic expansion coefficients is such that $0 < \delta_r < 1$, then $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}_r$ are spectrally equivalent

$$1 - \delta_r \leq \lambda(\tilde{\mathcal{S}}_r^{-1}\tilde{\mathcal{S}}) \leq 1 + \delta_r.$$

Proof. Consider the error matrix $\mathcal{E}_r = \tilde{\mathcal{S}} - \tilde{\mathcal{S}}_r$. Based on the expansion structure of the operator defined in (3.12), the error term consists of the neglected high-order terms in the stochastic expansion:

$$\mathcal{E}_r = \sum_{\ell=r+1}^{n_A} H_\ell \otimes \hat{A}_\ell,$$

where \hat{A}_ℓ represents the associated spatial coefficient matrices. We apply Lemma 4.2 by estimating the ratio $|\mathbf{v}^T \mathcal{E}_r \mathbf{v}|/|\mathbf{v}^T \tilde{\mathcal{S}}_r \mathbf{v}|$.

First, we bound the numerator using the triangle inequality and the submultiplicativity of the spectral norm:

$$|\mathbf{v}^T \mathcal{E}_r \mathbf{v}| = \left| \sum_{\ell=r+1}^{n_A} \mathbf{v}^T (H_\ell \otimes \hat{A}_\ell) \mathbf{v} \right| \leq \sum_{\ell=r+1}^{n_A} \left| \mathbf{v}^T (H_\ell \otimes \hat{A}_\ell) \mathbf{v} \right| \leq \sum_{\ell=r+1}^{n_A} \|H_\ell\|_2 \|\hat{A}_\ell\|_2 \|\mathbf{v}\|^2.$$

Next, we establish a lower bound for the denominator

$$\mathbf{v}^T \tilde{\mathcal{S}}_r \mathbf{v} \geq \mathbf{v}^T (H_1 \otimes \hat{A}_1) \mathbf{v} \geq \nu \|\mathbf{v}\|^2.$$

Combining these inequalities yields

$$\frac{|\mathbf{v}^T \mathcal{E}_r \mathbf{v}|}{|\mathbf{v}^T \tilde{\mathcal{S}}_r \mathbf{v}|} \leq \frac{\sum_{\ell=r+1}^{n_A} \|H_\ell\|_2 \|\hat{A}_\ell\|_2 \|\mathbf{v}\|^2}{\nu \|\mathbf{v}\|^2} = \frac{1}{\nu} \sum_{\ell=r+1}^{n_A} \|H_\ell\|_2 \|\hat{A}_\ell\|_2 = \delta_r.$$

Provided that the Karhunen-Loève expansion coefficients decay sufficiently fast (e.g., algebraically or exponentially), the tail sum approaches zero as $r \rightarrow n_A$. Thus, for sufficiently large r , we have $\delta_r < 1$. The result then follows directly from Lemma 4.2. \square

Finally, we proceed to establish the spectral equivalence between the truncated preconditioner $\tilde{\mathcal{S}}_r$ and the hierarchical symmetric block Gauss-Seidel approximation $\tilde{\mathcal{S}}_{\text{hGS-PINT}}$. To this end, following the idea from [6], we can rewrite it as $H_\ell = L_\ell + L_\ell^T$, $\ell = 2, 3, \dots, n_A$, and matrices L_ℓ have at most one nonzero entry per row and per column.

Now, define $X_1 = I_{n_t} \otimes H_1 \otimes \hat{A}_1$, $X_r = I_{n_t} \otimes \left(\sum_{\ell=2}^r L_\ell \otimes \hat{A}_\ell \right)$, $r = 2, 3, \dots, n_A$, so we know $\tilde{\mathcal{Z}}_{\text{hGS-PINT}} = (X_1 + X_r)X_1^{-1}(X_1 + X_r^T) = \tilde{\mathcal{Z}}_r + X_r X_1^{-1} X_r^T$, the Rayleigh quotient

$$\frac{\mathbf{v}^T \tilde{\mathcal{Z}}_{\text{hGS-PINT}} \mathbf{v}}{\mathbf{v}^T \tilde{\mathcal{Z}}_r \mathbf{v}} = 1 + \underbrace{\frac{\mathbf{v}^T X_r X_1^{-1} X_r^T \mathbf{v}}{\mathbf{v}^T (X_1 + X_r + X_r^T) \mathbf{v}}}_{\zeta}.$$

Introducing the change of variable $\mathbf{u} = X_1^{-\frac{1}{2}} \mathbf{v}$ and setting $Y = X_1^{-\frac{1}{2}} X_r X_1^{-\frac{1}{2}}$, we have

$$\zeta(\mathbf{u}) = \frac{\mathbf{u}^T Y Y^T \mathbf{u}}{\mathbf{u}^T (I + Y + Y^T) \mathbf{u}},$$

where $Y = I_{n_t} \otimes \left(\sum_{\ell=2}^r L_\ell \otimes \mathfrak{A}_\ell \right)$, with the scaled coefficient matrices defined as $\mathfrak{A}_\ell = \hat{A}_1^{-\frac{1}{2}} \hat{A}_\ell \hat{A}_1^{-\frac{1}{2}}$. Consequently,

$$\zeta(\mathbf{u}) \leq \max_{\mathbf{u} \neq \mathbf{0}} \frac{\mathbf{u}^T Y Y^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \cdot \max_{\mathbf{u} \neq \mathbf{0}} \frac{\mathbf{u}^T \mathbf{u}}{\mathbf{u}^T (I + Y + Y^T) \mathbf{u}} = \frac{\sigma_{\max}^2(Y)}{\lambda_{\min}(I + Y + Y^T)}.$$

The following result holds.

LEMMA 4.10. *Define*

$$\Delta_r := \sum_{\ell=2}^r \|H_\ell\|_2 \rho_\ell, \quad \rho_\ell := \|\mathfrak{A}_\ell\|_2,$$

and let

$$Y := X_1^{-1/2} X_r X_1^{-1/2} = I_{n_t} \otimes \left(\sum_{\ell=2}^r L_\ell \otimes \mathfrak{A}_\ell \right), \quad E := Y + Y^T = I_{n_t} \otimes \left(\sum_{\ell=2}^r H_\ell \otimes \mathfrak{A}_\ell \right),$$

with $H_\ell = L_\ell + L_\ell^T$. Then $\lambda_{\min}(I + Y + Y^T) \geq 1 - \Delta_r$, $\rho_\ell \leq \frac{\|\kappa_\ell(x)\|_{L^\infty(\Omega)}}{\mathbb{k}_{\min}}$, and

$$\sigma_{\max}(Y) \leq \sum_{\ell=2}^r \|L_\ell\|_2 \rho_\ell \leq \Delta_r.$$

Proof. Using

$$I + Y + Y^T = X_1^{-1/2} \left(X_1 + I_{n_t} \otimes \sum_{\ell=2}^r H_\ell \otimes \hat{A}_\ell \right) X_1^{-1/2} = X_1^{-1/2} \bar{\mathcal{Z}}_r X_1^{-1/2},$$

the spectrum of $I + Y + Y^T$ coincides with the generalized spectrum of the pair $(\bar{\mathcal{Z}}_r, X_1)$. Since E is symmetric, its spectral norm satisfies $\|E\|_2 = |\lambda_{\max}(E)| \geq |\lambda_i(E)|$ for any i . In particular, $\lambda_{\min}(E) \geq -\|E\|_2$, and therefore $\lambda_{\min}(I + E) = 1 + \lambda_{\min}(E) \geq 1 - \|E\|_2$. By the Kronecker product norm rule and the triangle inequality,

$$\|E\|_2 = \left\| I_{n_t} \otimes \left(\sum_{\ell=2}^r H_\ell \otimes \mathfrak{A}_\ell \right) \right\|_2 \leq \sum_{\ell=2}^r \|H_\ell\|_2 \|\mathfrak{A}_\ell\|_2 = \sum_{\ell=2}^r \|H_\ell\|_2 \rho_\ell = \Delta_r,$$

which yields $\lambda_{\min}(I + Y + Y^T) \geq 1 - \Delta_r$.

Similarly,

$$\|Y\|_2 = \left\| I_{n_t} \otimes \sum_{\ell=2}^r L_\ell \otimes \mathfrak{A}_\ell \right\|_2 \leq \sum_{\ell=2}^r \|L_\ell\|_2 \rho_\ell.$$

Since $H_\ell = L_\ell + L_\ell^T$ and each L_ℓ has at most one nonzero per row and per column, we have $\|L_\ell\|_2 \leq \|H_\ell\|_2$. Hence $\sigma_{\max}(Y) = \|Y\|_2 \leq \Delta_r$.

For the explicit bound on ρ_ℓ , recall that $\rho_\ell = \|\mathfrak{A}_\ell\|_2$ is the maximum eigenvalue of the generalized eigenvalue problem $A_\ell \mathbf{v} = \lambda \hat{A}_1 \mathbf{v}$, where \hat{A}_1 defined in (3.16). In terms of the associated finite element function $v_h = \sum_{j=1}^{n_h} v_j \phi_j$, the Rayleigh quotient is given by

$$\frac{\mathbf{v}^T A_\ell \mathbf{v}}{\mathbf{v}^T \hat{A}_1 \mathbf{v}} = \frac{\int_{\Omega} \kappa_\ell |\nabla v_h|^2 dx}{\int_{\Omega} \left((1 + \tau \sqrt{\frac{1+\gamma}{\beta}}) |v_h|^2 + \tau \kappa_0 |\nabla v_h|^2 \right) dx}.$$

Since the mass term is non-negative, we can bound this ratio by neglecting the L^2 -term in the denominator:

$$\frac{\mathbf{v}^T A_\ell \mathbf{v}}{\mathbf{v}^T \hat{A}_1 \mathbf{v}} \leq \frac{\int_{\Omega} \kappa_\ell |\nabla v_h|^2 dx}{\tau \int_{\Omega} \kappa_0 |\nabla v_h|^2 dx} \leq \frac{1}{\tau} \frac{\|\kappa_\ell(x)\|_{L^\infty(\Omega)}}{\mathbb{k}_{\min}},$$

and taking the supremum gives the stated bound on ρ_ℓ ; substituting it into $\|Y\|_2$ yields the last inequality. \square

Observe from above that, with $\Delta_r < 1$,

$$\zeta(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{Y} \mathbf{Y}^T \mathbf{v}}{\mathbf{v}^T (\mathbf{I} + \mathbf{Y} + \mathbf{Y}^T) \mathbf{v}} \leq \frac{\sigma_{\max}^2(\mathbf{Y})}{\lambda_{\min}(\mathbf{I} + \mathbf{Y} + \mathbf{Y}^T)} \leq \frac{\Delta_r^2}{1 - \Delta_r}.$$

Consequently, we have the bound for the Rayleigh quotient of the factors:

$$\frac{\mathbf{v}^T (\tilde{\mathcal{Z}}_{\text{hGS-PINT}} - \tilde{\mathcal{Z}}_r) \mathbf{v}}{\mathbf{v}^T \tilde{\mathcal{Z}}_r \mathbf{v}} = \zeta(\mathbf{v}) \leq \frac{\Delta_r^2}{1 - \Delta_r}.$$

To derive the bound for the preconditioner $\tilde{\mathcal{S}}$, we utilize Lemma 4.3. Let $Q = (D \otimes \mathcal{M})^{1/2}$. The generalized eigenvalues of $(\tilde{\mathcal{S}}_{\text{hGS-PINT}}, \tilde{\mathcal{S}}_r)$ are identical to those of the pair

$$\left((Q \tilde{\mathcal{Z}}_{\text{hGS-PINT}} Q)^2, (Q \tilde{\mathcal{Z}}_r Q)^2 \right).$$

Since the eigenvalues of the squared operators are simply the squares of the eigenvalues of the base operators (for these symmetric positive definite factors), the spectral bound for the preconditioner is the square of the bound for the factors. Thus, we conclude:

$$1 \leq \lambda(\tilde{\mathcal{S}}_r^{-1} \tilde{\mathcal{S}}_{\text{hGS-PINT}}) \leq \left(1 + \frac{\Delta_r^2}{1 - \Delta_r} \right)^2.$$

REMARK 1 (Summary of Spectral Equivalence for Time-Dependent Case). *Under the assumptions of Theorem 4.6, Lemma 4.7, Theorem 4.9, and Lemma 4.10, the hierarchical Gauss-Seidel PINT preconditioner $\tilde{\mathcal{S}}_{\text{hGS-PINT}}$ is spectrally equivalent to the exact Schur complement $\tilde{\mathcal{S}}_{\text{exact}}$. The spectral equivalence is established through the following chain of approximations:*

$$\tilde{\mathcal{S}}_{\text{exact}} \sim \tilde{\mathcal{S}} \sim \tilde{\mathcal{S}} \sim \tilde{\mathcal{S}}_r \sim \tilde{\mathcal{S}}_{\text{hGS-PINT}}.$$

In particular, provided that the truncation rank r is sufficiently large (to satisfy $\delta_r < 1$) and the mass matrix M is well-conditioned (e.g., via mass lumping), the eigenvalues of the preconditioned system $\tilde{\mathcal{S}}_{\text{hGS-PINT}}^{-1} \tilde{\mathcal{S}}_{\text{exact}}$ are contained in a fixed interval independent of the spatial mesh size h and the stochastic discretization parameters m_ξ and p (assuming sufficient decay of the expansion coefficients), ensuring mesh robustness in the fixed time step regime.

REMARK 2 (Steady-State Case). *Since the steady-state optimal control problem corresponds to the special case $n_t = 1$ of the time-dependent formulation, all preceding results apply directly with the simplified notation. The spectral equivalence chain for the steady-state Schur complement preconditioner is*

$$\mathcal{S}_{\text{exact}} \sim \mathcal{S} \sim \mathcal{S}_r \sim \mathcal{S}_{\text{hGS}},$$

where $\mathcal{S}_{\text{exact}}$ is defined in (2.12), \mathcal{S} in (2.13), and

$$\mathcal{S}_r = \mathcal{Z}_r \mathcal{M}_\gamma^{-1} \mathcal{Z}_r^T, \quad \mathcal{Z}_r = \sum_{\ell=1}^r H_\ell \otimes \tilde{A}_\ell,$$

with $\tilde{A}_1 = A_1 + \sqrt{\frac{1+\gamma}{\beta}} M$ and $\tilde{A}_\ell = A_\ell$ for $\ell = 2, \dots, r$. When $r = 1$, \mathcal{S}_r reduces to the mean-based preconditioner, and when $r = n_A$, it recovers the full operator \mathcal{S} . Finally, \mathcal{S}_{hGS} represents the computationally feasible approximation of \mathcal{S}_r in which the linear systems $\mathcal{Z}_r \mathbf{x} = \mathbf{b}$ are solved approximately via the hierarchical Gauss-Seidel method (Algorithm 2.1), as implemented in Algorithm 2.2–2.3.

5. Numerical experiments. This section validates the theoretical findings of Sections 2.1–3 through comprehensive numerical experiments. We pursue two primary objectives: (i) verifying the mesh-robust for fixed time step and spectral bounds established in the preceding sections, and (ii) demonstrating the computational efficiency of the proposed hierarchical Gauss-Seidel (hGS) preconditioner across varying truncation strategies. Experiments are presented for both steady-state problems (Section 5.1) and time-dependent problems (Section 5.2). The numerical experiments were performed on a system running AlmaLinux-9 with 40GB RAM, and the proposed algorithms were implemented using MATLAB 23.2.

The random coefficients $\mathbb{k}(x, \xi)$ in the problem (2.2) are constructed as a finite expansion (2.4), where the spatial modes $\kappa_i(x)$ and weights θ_i are eigenpairs of the covariance function

$$C_f(x, y) = \sigma_{\mathbb{k}}^2 \exp\left(-\frac{|x_1 - y_1|}{\ell_1} - \frac{|x_2 - y_2|}{\ell_2}\right) \quad \forall (x, y) \in [-1, 1]^2.$$

We set the correlation lengths as $\ell_1 = \ell_2 = 1$, the mean $\kappa_1(x) \equiv 1$. For the gPC setting in (2.5), we consider a log-normal random field parameterized by independent Gaussian random variables, for which corresponding Hermite polynomials are employed as the basis. The total number of basis functions is given by $n_{\xi} = \binom{m_{\xi} + p}{m_{\xi}}$, where m_{ξ} is the number of random variables and p is the polynomial degree. For instance, the case $(m_{\xi}, p) = (3, 3)$ yields $n_{\xi} = \binom{6}{3} = 20$. This problem has been extensively studied in [20]. Also, we used $\gamma = 1$ in both cases, which means we only consider the case with standard deviation. To discretize the spatial domain, we implemented our code based on IFISS 3.7 [24], using \mathbf{Q}_1 approximation. For temporal discretization, we apply the all-at-once technique proposed in [21] and set the terminal time as $T = 1$. In all numerical experiments, the spatial mesh size h and the time step τ are chosen as 2^{-i} , with $i = 4, 5, 6, 7$. We solve the linear systems (2.8) and (3.7) using the preconditioners given by (2.11) and (3.8), respectively, employing the flexible GMRES method (without restarting) [23]. The stopping criterion is defined in terms of the relative residual $\|r_k\|/\|b\|$, with thresholds 10^{-8} for the steady-state experiments and 10^{-6} or 10^{-4} for the time-dependent runs, where r_k denotes the residual at iteration k and b is the right-hand side vector. For notational consistency, we set $n_{\tau} \equiv r$ throughout the remainder of the paper. To assess the effectiveness of the hierarchical preconditioning strategy, we systematically compare three truncation settings for the (3,3)-block preconditioner: $r = 1$ (mean-based approximation), $r = m_{\xi} + 1$ (hGS truncated at the first-order stochastic terms), and $r = n_A$ (full expansion retaining all h_{ijk} coefficients). For each configuration, we evaluate the efficiency of the hierarchical Gauss-Seidel method by comparing both the iteration counts and the computational times (in seconds).

We consider homogeneous Dirichlet conditions (i.e., $g(x) = 0$), corresponding to Example 2 in [9, Chapter 5]. This example is defined on a square domain $\mathcal{D} = [-1, 1]^2$ with a discontinuous target function

$$(5.1) \quad y_d = \begin{cases} 1 & \text{in } \Omega_1 := [-1, 0]^2, \\ 0 & \text{in } \mathcal{D} \setminus \Omega_1, \end{cases}$$

which represents inconsistent Dirichlet boundary data since the target state $y_d = 1$ differs from the required boundary condition $y = 0$ on $\partial\mathcal{D} \cap \partial\Omega_1$.

We subsequently present numerical experiments for both steady-state and time dependent problems to illustrate and verify the efficiency of our proposed hGS method.

Here are some details about the implementation for preconditioners (2.11), (3.8). Since time-dependent problems can be seen as a series of steady-state problems, and also because of the diagonal structure of matrix D and matrix I_{n_t} , we can just focus on the steady-state preconditioner.

The practical implementation of the preconditioner \mathcal{P} in (2.11) involves different strategies for its constituent blocks. For the (1,1) and (2,2) blocks, which are based on Kronecker products involving the mass matrix M , applying their inverses requires solving linear systems with M . These solves are handled efficiently by either a direct Cholesky decomposition or the iterative Chebyshev semi-iteration method [12, 28].

For the more complex (3,3) block, which represents the approximate Schur complement \mathcal{S} , we employ an outer iterative scheme. Specifically, we use the Preconditioned Richardson method, outlined in Algorithm 5.1 [7, Chapter 7], where the core of our proposal—the hierarchical Gauss-Seidel (hGS) method from Algorithm 2.1 serves as the preconditioner for each Richardson step.

Algorithm 5.1 Preconditioned Richardson iteration with hGS preconditioner

- 1: Given matrix \mathcal{Z} , vector b , and initial guess $x^{(1)}$.
 - 2: $r^{(1)} = b - \mathcal{Z}x^{(1)}$. (initial residual)
 - 3: **for** $k = 1, 2, \dots, N$ **do**
 - 4: Solve $\mathcal{Z}z^{(k)} = r^{(k)}$. (apply Algorithm 2.1)
 - 5: $x^{(k+1)} = x^{(k)} + z^{(k)}$ (update solution)
 - 6: $r^{(k+1)} = b - \mathcal{Z}x^{(k+1)}$. (update residual)
 - 7: **end for**
 - 8: **Return** $x^{(N+1)}$
-

5.1. Steady-state case. This subsection focuses on the steady-state optimal control problem (2.1). We examine the performance of the proposed preconditioner (2.11) under systematic variations in: (i) the variance $\sigma_{\mathbf{k}}$ (Tables 5.1–5.3), (ii) the regularization parameter β (Table 5.4), and (iii) the spatial and stochastic discretization levels. For each configuration, we compare three solvers for the (1,1) and (2,2) blocks—Chebyshev semi-iteration with 5 or 10 steps and direct Cholesky factorization—combined with the three truncation strategies for the (3,3)-block described above. For the (3,3)-block, we apply the Richardson iteration (Algorithm 5.1) with $N = 1$, i.e., one application of the hGS preconditioner per outer GMRES iteration. All tests use a fixed tolerance of 10^{-8} .

Next, by fixing the parameter β , we perform further tests summarized in Tables 5.1–5.3, employing different step settings for the Chebyshev smoother and the Cholesky decomposition for blocks (1,1) and (2,2), as well as various truncation strategies for block (3,3) within the hGS method. The numerical experiments were conducted using multiple mesh sizes and stochastic parameter configurations, with a solver tolerance set to 10^{-8} .

Tables 5.1–5.3 present results for $\beta = 10^{-4}$ with $\sigma_{\mathbf{k}} \in \{0.01, 0.1, 0.4\}$, covering a range from near-deterministic to highly stochastic regimes. Tables 5.1–5.3 demonstrate three key theoretical properties. First, regarding *mesh independence*, iteration counts grow sub-linearly with spatial refinement for fixed stochastic dimension n_{ξ} , consistent with the spectral bounds established in Section 2.1. Second, concerning *truncation efficiency*, the $n_{\tau} = m_{\xi} + 1$ strategy achieves iteration counts comparable to the full expansion ($n_{\tau} = n_A$) while avoiding the computational overhead of summing over all h_{ijk} coefficients, thereby validating the hierarchical approximation framework.

TABLE 5.1

Results showing the total number of iterations from low-rank preconditioned GMRES and the total CPU times (in seconds) using preconditioner with $\beta = 10^{-4}$, $\sigma = 0.01$, and selected spatial (n_h) and stochastic (n_ξ) degrees of freedom

	# iter(t)			# iter(t)			# iter(t)			# iter(t)		
$n_\xi \backslash n_h$	$289(h = \frac{1}{2^7})$			$1089(h = \frac{1}{2^8})$			$4225(h = \frac{1}{2^9})$			$16641(h = \frac{1}{2^{10}})$		
n_τ	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A
$\sigma_a = 0.01$												
Chebyshev-5+hGS-1												
20	32(5.7)	32(4.0)	32(5.0)	36(9.2)	35(10.0)	35(15.6)	36(23.2)	27(26.9)	27(46.5)	35(173.4)	33(144.9)	27(194.8)
70	32(129.1)	32(129.5)	32(182.9)	36(243.4)	35(239.1)	35(409.2)	36(752.3)	34(724.9)	34(1237.8)	35(2585.4)	33(2524.6)	33(5253.9)
84	32(317.5)	32(318.3)	32(403.0)	36(508.9)	35(498.2)	35(601.3)	36(1059.3)	34(864.8)	34(1632.5)	35(3099.0)	33(3798.1)	33(5630.2)
Chebyshev-10+hGS-1												
20	26(2.8)	26(2.6)	26(4.4)	30(6.9)	29(7.1)	29(12.0)	31(20.2)	31(21.1)	31(37.6)	32(145.1)	31(130.3)	31(154.5)
70	26(111.7)	26(112.2)	26(153.7)	30(191.6)	29(189.0)	29(323.8)	32(550.7)	31(383.1)	31(830.2)	33(1745.9)	31(2273.0)	31(3341.3)
84	26(263.9)	26(266.1)	26(329.2)	30(457.7)	29(432.2)	29(504.1)	32(959.1)	31(800.4)	31(1515.3)	33(3015.1)	31(3643.9)	31(5318.7)
Cholesky+hGS-1												
20	25(3.5)	25(4.0)	25(4.3)	29(7.4)	27(6.7)	27(10.9)	29(19.1)	29(26.5)	29(44.5)	31(143.6)	29(162.7)	29(182.9)
70	25(104.2)	25(107.6)	25(144.3)	29(177.3)	27(168.9)	27(283.9)	29(419.0)	29(480.8)	29(937.1)	31(2390.7)	29(2046.0)	29(3299.9)
84	25(245.4)	25(247.4)	25(316.8)	29(429.9)	27(393.6)	27(675.5)	29(831.9)	29(1008.3)	29(1756.8)	31(3563.3)	29(3393.6)	29(6436.3)

TABLE 5.2

Results showing the total number of iterations from low-rank preconditioned GMRES and the total CPU times (in seconds) using preconditioner with $\beta = 10^{-4}$, $\sigma = 0.1$, and selected spatial (n_h) and stochastic (n_ξ) degrees of freedom

	# iter(t)			# iter(t)			# iter(t)			# iter(t)		
$n_\xi \backslash n_h$	$289(h = \frac{1}{2^7})$			$1089(h = \frac{1}{2^8})$			$4225(h = \frac{1}{2^9})$			$16641(h = \frac{1}{2^{10}})$		
n_τ	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A
$\sigma_a = 0.1$												
Chebyshev-5+hGS-1												
20	39(3.6)	33(3.3)	33(3.7)	35(11.1)	35(8.7)	29(14.2)	34(29.7)	36(40.6)	35(42.3)	34(162.3)	28(84.0)	34(191.2)
70	41(199.9)	34(173.7)	33(136.4)	45(218.5)	35(190.5)	35(299.8)	46(841.0)	36(670.3)	35(903.4)	45(2093.8)	35(2917.1)	34(4099.5)
84	41(460.9)	34(238.3)	33(325.3)	45(475.0)	35(529.5)	35(610.9)	44(1332.7)	36(1028.1)	35(1708.7)	44(4940.4)	35(4008.7)	34(5843.5)
Chebyshev-10+hGS-1												
20	34(2.8)	26(2.2)	26(3.1)	38(7.1)	30(7.0)	30(10.5)	40(25.7)	31(32.8)	31(37.8)	42(108.7)	31(86.2)	31(152.0)
70	36(191.4)	26(139.4)	26(114.9)	40(186.6)	30(234.7)	30(264.6)	42(488.8)	31(518.1)	31(894.2)	44(1746.1)	32(1518.3)	31(3608.8)
84	35(411.8)	26(193.3)	26(283.1)	40(436.9)	30(354.9)	30(558.3)	42(1221.2)	31(970.6)	31(1529.2)	43(4988.0)	32(3801.3)	31(5417.9)
Cholesky+hGS-1												
20	33(2.6)	25(2.2)	25(3.2)	37(7.9)	27(6.9)	27(11.1)	39(28.4)	29(34.0)	29(34.8)	39(129.7)	29(103.2)	29(150.1)
70	35(177.5)	25(127.6)	25(177.2)	39(282.8)	29(216.6)	27(320.4)	41(603.6)	29(581.4)	29(818.1)	41(1928.1)	29(1582.8)	29(3003.8)
84	35(401.1)	25(332.4)	25(407.1)	39(657.8)	29(445.5)	27(665.0)	41(1156.5)	29(869.2)	29(1706.1)	41(4696.6)	29(2848.4)	29(22402.2)

Third, regarding *smoother comparison*, the 5-step Chebyshev semi-iteration balances convergence rate and per-iteration cost more effectively than either the 10-step variant or direct Cholesky factorization. Across all configurations, the $n_\tau = m_\xi + 1$ truncation consistently delivers performance intermediate between the mean-based approximation ($n_\tau = 1$) and the full expansion, confirming the practical value of the proposed hierarchical preconditioning strategy.

Table 5.4 examines the sensitivity to the regularization parameter β , which balances the tracking term and control cost in the objective functional (2.1). As β decreases from 10^{-2} to 10^{-5} , the optimization problem becomes increasingly dominated by the tracking term. The iteration counts remain remarkably stable across this range, demonstrating that the hierarchical preconditioner effectively handles varying parameter regimes without requiring problem-specific tuning. The $n_\tau = m_\xi + 1$ truncation consistently performs comparably to the full expansion while maintaining reduced computational cost.

5.2. Time-dependent case. This subsection evaluates the all-at-once preconditioner (3.8) for time-dependent optimal control problems. The discretization results in KKT systems of dimension $n_t \times n_\xi \times n_h$, where n_t denotes the number of time

TABLE 5.3

Results showing the total number of iterations from low-rank preconditioned GMRES and the total CPU times (in seconds) using preconditioner with $\beta = 10^{-4}$, $\sigma = 0.4$, and selected spatial (n_h) and stochastic (n_ξ) degrees of freedom

	# iter(t)			# iter(t)			# iter(t)			# iter(t)		
$n_h \backslash n_\xi$	$289(h = \frac{1}{2^7})$			$1089(h = \frac{1}{2^5})$			$4225(h = \frac{1}{2^3})$			$16641(h = \frac{1}{2^1})$		
n_τ	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A
$\sigma_a = 0.4$												
Chebyshev-5+hGS-1												
20	59(5.7)	36(3.6)	27(4.5)	84(13.3)	39(10.1)	29(13.2)	88(50.9)	38(37.8)	36(60.4)	88(398.2)	38(163.3)	28(203.0)
70	100(275.7)	38(198.3)	34(140.1)	112(498.4)	42(299.3)	37(316.8)	118(1792.6)	41(697.2)	36(929.6)	117(7712.6)	41(2655.2)	36(3742.7)
84	86(720.2)	36(414.0)	34(463.0)	96(986.0)	40(596.4)	37(624.8)	102(3051.9)	40(1452.6)	37(1838.1)	102(9946.2)	41(4576.9)	36(7289.7)
Chebyshev-10+hGS-1												
20	69(6.4)	30(2.8)	28(3.3)	77(13.9)	33(8.2)	30(9.9)	81(50.5)	35(34.4)	31(38.1)	83(208.6)	36(148.7)	33(162.3)
70	92(278.4)	32(96.5)	28(125.9)	105(508.4)	36(175.3)	30(269.3)	110(1318.5)	37(781.9)	32(857.9)	110(5212.5)	38(2920.1)	33(4546.2)
84	79(931.3)	30(355.0)	28(411.9)	89(1373.9)	34(529.2)	30(531.0)	95(3528.4)	36(1337.8)	32(1576.4)	97(11001.9)	38(4413.7)	33(5769.8)
Cholesky+hGS-1												
20	69(6.8)	29(3.0)	27(4.0)	77(17.5)	31(8.2)	29(11.7)	81(57.5)	33(35.2)	29(50.3)	83(279.8)	33(180.9)	31(181.8)
70	31(159.1)	31(156.3)	29(350.5)	105(730.2)	33(238.6)	29(331.6)	109(1903.1)	35(726.1)	31(953.3)	109(7600.0)	35(2737.8)	31(3350.7)
84	33(753.1)	33(353.2)	27(425.5)	89(1325.8)	33(493.8)	29(634.8)	95(3122.6)	35(1055.9)	31(1837.7)	97(10923.5)	35(4023.8)	31(6429.3)

TABLE 5.4

Results using the preconditioner with $m_\xi=3$, $p=3$, $\sigma_k = 0.2$, $\beta \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ and $n_h = 1089(h = \frac{1}{2^5})$.

	# iter(t)			# iter(t)			# iter(t)		
n_ξ	20			70			84		
n_τ	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A
Chebyshev-5+hGS-1									
$\beta = 10^{-2}$	52(10.0)	32(7.9)	30(13.5)	60(272.3)	32(149.2)	30(363.0)	56(596.9)	32(338.9)	30(626.9)
$\beta = 10^{-3}$	52(10.3)	34(8.3)	34(15.5)	60(273.8)	35(161.7)	34(411.4)	56(594.0)	34(360.7)	34(709.7)
$\beta = 10^{-4}$	43(11.7)	37(7.6)	35(15.6)	62(279.7)	37(172.1)	35(299.2)	58(629.0)	37(393.4)	35(731.8)
$\beta = 10^{-5}$	41(11.2)	38(7.9)	37(17.0)	60(270.6)	38(175.9)	37(317.9)	56(589.0)	38(413.3)	38(795.5)
Chebyshev-10+hGS-1									
$\beta = 10^{-2}$	48(8.7)	30(6.1)	29(11.9)	56(420.8)	29(147.9)	29(359.7)	52(577.6)	34(370.8)	24(653.1)
$\beta = 10^{-3}$	48(8.8)	31(6.4)	30(12.6)	56(432.3)	31(151.6)	30(406.9)	52(554.8)	31(339.6)	30(675.8)
$\beta = 10^{-4}$	48(8.8)	30(7.1)	30(19.4)	56(419.0)	31(151.2)	30(390.6)	52(575.0)	31(339.6)	30(676.1)
$\beta = 10^{-5}$	47(8.7)	30(7.0)	30(13.2)	54(406.4)	30(146.0)	30(383.9)	51(554.0)	30(333.0)	30(675.1)
Chol+hGS-1									
$\beta = 10^{-2}$	47(12.4)	29(6.9)	27(11.9)	55(380.1)	29(214.4)	27(354.4)	51(975.8)	29(483.9)	27(595.1)
$\beta = 10^{-3}$	47(12.0)	29(6.4)	27(11.9)	55(396.0)	29(215.9)	27(364.5)	51(876.9)	29(455.6)	27(596.0)
$\beta = 10^{-4}$	47(10.4)	29(7.4)	27(12.2)	55(387.9)	29(216.0)	27(343.9)	53(847.9)	29(432.8)	27(595.5)
$\beta = 10^{-5}$	45(10.0)	29(6.8)	27(12.0)	53(369.6)	29(205.5)	27(359.3)	51(829.3)	29(438.9)	27(594.1)

steps. We investigate the scalability with respect to: (i) mesh refinement (Table 5.5), (ii) regularization parameter β (Table 5.6), (iii) variance σ_k (Table 5.7), (iv) time-step size discretization τ (Table 5.8), and (v) stochastic dimension (m_ξ, p) (Table 5.9). Based on the steady-state findings, we employ the Chebyshev-5+hGS-1 configuration unless otherwise noted, reporting results for both stringent (10^{-6}) and moderate (10^{-4}) tolerances to illustrate practical convergence behavior. As in the steady-state case, we use $N = 1$ in the Richardson iteration (Algorithm 5.1) for the (3,3)-block.

From our observations in the steady-state problem, a combination of a 5-step Chebyshev smoother with one step of our hGS method achieves a good balance between the computational cost of matrix operations and GMRES iterations; thus, we typically adopt this combination when testing time-dependent cases as well.

As indicated in Table 5.5, the 5-step Chebyshev smoother yields consistent iteration counts compared to either the 10-step smoother or direct Cholesky decomposition. As the spatial discretization is refined from $h = \frac{1}{2^3}$ to $h = \frac{1}{2^6}$, representing a growth

from 116,640 to 6,084,000 DoF, the iteration count for $n_\tau = m_\xi + 1$ exhibits sub-linear growth consistent with the near mesh-independence predicted by the spectral theory in Section 4. The $n_\tau = m_\xi + 1$ truncation achieves iteration counts comparable to the full expansion ($n_\tau = n_A$) while reducing the cost of assembling and applying the (3,3)-block preconditioner—a trade-off that becomes increasingly favorable as the problem dimension grows.

TABLE 5.5

Results using hGS method with different truncation settings n_τ with the time-dependent model for different stopping tolerance and mesh size with $m_\xi=3$, $p=3$, $\sigma_k=0.2$, $\beta=10^{-4}$ and number of time steps=8 ($\tau = \frac{1}{2^3}$).

h	n_τ DoF	tol= 10^{-6}			tol= 10^{-4}		
		1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A
$\frac{1}{2^3}$	116,640	57(11.9)	39(8.8)	39(15.7)	41(6.8)	31(5.7)	29(8.8)
$\frac{1}{2^4}$	416,160	75(39.3)	47(29.6)	45(39.6)	55(21.7)	35(14.9)	35(24.7)
$\frac{1}{2^5}$	1,568,160	83(150.9)	53(103.3)	51(158.6)	66(90.6)	43(58.8)	39(92.1)
$\frac{1}{2^6}$	6,084,000	84(570.8)	55(383.3)	53(585.1)	68(401.4)	42(252.2)	42(424.1)

Table 5.6 examines four orders of magnitude for β , ranging from 10^{-2} (control-dominant) to 10^{-8} (tracking-dominant). The mean-based preconditioner ($n_\tau = 1$) exhibits strong dependence on β , with iteration counts decreasing as β decreases (since smaller β yields problems dominated by the simpler tracking term). In contrast, the $n_\tau = m_\xi + 1$ truncation maintains stable iteration counts across all tested values, demonstrating that the hGS preconditioner automatically adapts to the problem structure without manual parameter tuning. This robustness confirms the applicability of the theoretical framework across diverse parameter regimes.

TABLE 5.6

Results using hGS method with different truncations setting n_τ with the time-dependent model for different stopping tolerance and β , mesh size $h = \frac{1}{2^5}$, $m_\xi=3$, $p=3$, $\sigma_k=0.2$, and number of time steps=8 ($\tau = \frac{1}{2^3}$), which results in 1,568,160 DoF.

β	n_τ	tol= 10^{-6}			tol= 10^{-4}		
		1	$m_\xi+1$	n_A	1	$m_\xi+1$	n_A
10^{-2}		107(206.1)	69(136.3)	69(214.8)	74(102.0)	48(70.4)	47(112.6)
10^{-3}		116(164.4)	74(109.4)	74(174.6)	66(90.6)	42(60.7)	42(100.0)
10^{-6}		97(137.0)	77(115.6)	77(182.5)	57(77.4)	45(65.3)	44(103.7)

Table 5.7 explores the range $\sigma_k \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.4\}$, spanning from nearly deterministic to highly uncertain regimes. The mean-based preconditioner ($n_\tau = 1$) exhibits significant degradation as uncertainty increases, whereas the $n_\tau = m_\xi + 1$ truncation maintains stable iteration counts across the entire range. Especially when σ_k increases from 0.2 to 0.4, the mean-based preconditioner performs poorly with a large number of iterations, but the hGS method maintains robust performance. This robustness confirms that the hierarchical preconditioner effectively captures the essential stochastic structure without requiring full expansion of all coupling coefficients. Table 5.8 investigates the all-at-once system scalability by varying n_t from 4 to 256 (time steps $\tau \in \{1/4, 1/16, 1/64, 1/256\}$), corresponding to total system sizes ranging from 784,080 to over 12.5 million DoF. As the temporal resolution increases, the coupled space-time-stochastic system grows proportionally, yet the $n_\tau = m_\xi + 1$ truncation maintains sub-linear iteration growth relative to system size. The computational time

TABLE 5.7

Results using hGS method with different truncations setting n_τ with the time-dependent model for different stopping tolerance and $\sigma_{\mathbf{k}}$, $\beta = 10^{-4}$, mesh size $n_h = 1089$, $m_\xi = 3$, $p = 3$, and number of time steps = 8 ($\tau = \frac{1}{2^3}$), which results in 1,568,160 DoF.

		tol= 10^{-6}			tol= 10^{-4}		
		n_τ	$m_\xi + 1$	n_A	n_τ	$m_\xi + 1$	n_A
$\sigma_{\mathbf{k}}$	0.01	51(77.7)	51(85.3)	51(133.9)	43(68.6)	41(72.7)	41(112.4)
	0.02	53(80.8)	51(82.2)	51(130.0)	43(70.2)	41(71.1)	41(113.7)
	0.05	59(90.2)	51(82.1)	51(130.9)	47(78.3)	41(72.6)	41(113.1)
	0.1	65(99.2)	51(82.4)	51(130.4)	53(86.2)	42(70.6)	41(116.5)
	0.2	83(150.9)	53(103.3)	51(158.6)	66(90.6)	43(58.8)	39(92.1)
	0.4	129(200.9)	57(92.3)	51(130.5)	100(179.3)	47(82.3)	43(119.2)

scales approximately linearly with DoF, confirming the efficiency of the all-at-once preconditioner for massively coupled systems.

TABLE 5.8

Results using hGS method with different truncations setting n_τ with the time-dependent model for different stopping tolerance and $\sigma_{\mathbf{k}}$, $\beta = 10^{-4}$, mesh size $n_h = 1089$, $m_\xi = 3$, $p = 3$, and number of time steps = 8 ($\tau = \frac{1}{2^3}$).

		tol= 10^{-6}			tol= 10^{-4}			
		n_τ	$m_\xi + 1$	n_A	n_τ	$m_\xi + 1$	n_A	
τ	DoF							
	$1/2^2$	784,080	81(78.6)	51(51.1)	51(71.1)	66(53.9)	43(37.6)	42(57.2)
	$1/2^4$	3,136,320	85(238.2)	55(166.4)	53(258.7)	68(174.2)	44(124.3)	43(203.7)
	$1/2^6$	12,545,280	101(1029.8)	67(733.3)	65(1205.8)	78(840.5)	53(604.2)	52(1028.6)

Finally, Table 5.9 compares three gPC configurations: $(m_\xi, p) \in \{(3, 3), (4, 4), (6, 3)\}$, yielding $n_\xi \in \{20, 70, 84\}$ basis functions. Table 5.9 varies the stochastic discretization parameters (m_ξ, p) , exploring both the number of random variables and polynomial order. As n_ξ increases from 20 to 84, the iteration count for $n_\tau = m_\xi + 1$ grows modestly, demonstrating near-independence from the stochastic discretization level. This behavior confirms the effectiveness of the hierarchical truncation strategy in maintaining spectral properties across varying gPC expansion settings.

TABLE 5.9

Results using hGS method with different truncations setting n_τ with the time-dependent model for different stopping tolerance and stochastic setting, $\beta = 10^{-4}$, mesh size $n_h = 1089$, $\sigma_{\mathbf{k}} = 0.2$, and number of time steps = 8 ($\tau = \frac{1}{2^3}$).

		tol= 10^{-6}			tol= 10^{-4}			
		n_τ	$m_\xi + 1$	n_A	n_τ	$m_\xi + 1$	n_A	
n_ξ	DoF							
	20	1,568,160	83(150.9)	53(103.3)	51(158.6)	66(90.6)	43(58.8)	39(92.1)
	70	5,488,560	95(2435.3)	53(1397.4)	52(2964.7)	74(1809.2)	43(1044.9)	43(2476.2)
	84	6,586,272	93(4092.4)	56(2670.3)	52(5207.4)	70(3059.4)	43(2107.7)	43(4413.4)

The time-dependent experiments establish that the proposed all-at-once preconditioner maintains robust performance across a wide range of problem parameters and discretization levels. Four key findings emerge from these results. First, regarding *near mesh-independence*, iteration growth remains sub-linear with spatial refinement (Table 5.5), consistent with the spectral bounds derived in Section 4. Second, concerning *parameter robustness*, the hGS method adapts automatically to varying β

(Table 5.6) and $\sigma_{\mathbf{k}}$ (Table 5.7) without manual tuning. Third, regarding *truncation efficiency*, the $n_{\tau} = m_{\xi} + 1$ strategy consistently delivers performance comparable to the full expansion at significantly reduced cost. These results demonstrate that the hierarchical preconditioning framework extends seamlessly from steady-state to time-dependent problems, providing a practical and theoretically-grounded solution for large-scale stochastic optimal control problems.

6. Conclusions. In the paper, we designed, analyzed, and implemented a novel hierarchical preconditioning strategy for large-scale stochastic optimal control problems. Our approach leverages a truncated stochastic expansion within a block-structured preconditioner for the Karush-Kuhn-Tucker (KKT) system, striking an effective balance between computational cost and preconditioning quality. Numerical results confirm that the proposed hGS method consistently outperforms both standard mean-based preconditioner and computationally intensive full-expansion preconditioner across a wide range of problem parameters.

A key contribution of this work is the extension of the truncated gPC framework to time-dependent problems. We developed and tested a tailored hGS preconditioner within an all-at-once discretization scheme, demonstrating the versatility and effectiveness of our approach for these challenging, large-scale scenarios. Comprehensive numerical experiments on benchmark problems have validated the robustness and numerical efficiency of the proposed algorithms.

REFERENCES

- [1] H. ANTIL, S. DOLGOV, AND A. ONWUNTA, *TTRisk: Tensor train decomposition algorithm for risk averse optimization*, Numer. Linear Algebra Appl., 30 (2023), p. 2481.
- [2] I. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.
- [3] P. BENNER, S. DOLGOV, A. ONWUNTA, AND M. STOLL, *Low-rank solvers for unsteady Stokes–Brinkman optimal control problem with random data*, Comput. Methods Appl. Mech. Engrg., 304 (2016), pp. 26–54.
- [4] P. BENNER, S. DOLGOV, A. ONWUNTA, AND M. STOLL, *Low-rank solution of an optimal control problem constrained by random Navier-Stokes equations*, Int. J. Numer. Methods Fluids, 92 (2020), pp. 1653–1678.
- [5] P. BENNER, A. ONWUNTA, AND M. STOLL, *Block-diagonal preconditioning for optimal control problems constrained by PDEs with uncertain inputs*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 491–518.
- [6] A. BESPALOV, D. LOGHIN, AND R. YOUNGNOI, *Truncation preconditioners for stochastic Galerkin finite element discretizations*, SIAM J. Sci. Comput., 43 (2021), pp. S92–S116.
- [7] A. K. BJÖRCK, *Numerical methods for least squares problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.
- [8] P. CHEN AND A. QUARTERONI, *Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraint*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 364–396.
- [9] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*, Oxford University Press, 2014.
- [10] R. GHANEM, *The nonlinear Gaussian spectrum of log-normal stochastic processes and variables*, J. Appl. Mech., 66 (1999), pp. 964–973.
- [11] R. G. GHANEM AND P. D. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag New York, Inc., New York, NY, USA, 1991.
- [12] G. H. GOLUB AND R. S. VARGA, *Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods*, Numer. Math., 3 (1961), pp. 157–168.
- [13] D. P. KOURI, *A multilevel stochastic collocation algorithm for optimization of PDEs with uncertain coefficients*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 55–81.
- [14] O. LE MAÎTRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification: With*

- Applications to Computational Fluid Dynamics*, Scientific Computation, Springer, 2010.
- [15] H.-C. LEE AND J. LEE, *A stochastic Galerkin method for stochastic control problems*, Commun. Comput. Phys., 14 (2013), pp. 77–106.
 - [16] G. J. LORD, C. E. POWELL, AND T. SHARDLOW, *An Introduction to Computational Stochastic PDEs*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2014.
 - [17] J. MARTÍNEZ-FRUTOS AND F. PERIAGO ESPARZA, *Optimal Control of PDEs under Uncertainty: An Introduction with Application to Optimal Shape Design of Structures*, Springer Briefs in Mathematics, Springer Cham, 2018.
 - [18] H. G. MATTHIES AND A. KEESE, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1295–1331.
 - [19] J. W. PEARSON, M. STOLL, AND A. J. WATHEN, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1126–1152.
 - [20] C. E. POWELL AND H. C. ELMAN, *Block-diagonal preconditioning for spectral stochastic finite-element systems*, IMA J. Numer. Anal., 29 (2009), pp. 350–375.
 - [21] T. REES, M. STOLL, AND A. WATHEN, *All-at-once preconditioning in PDE-constrained optimization*, Kybernetika, 46 (2010), pp. 341–360.
 - [22] E. ROSSEEL AND G. N. WELLS, *Optimal control with stochastic PDE constraints and uncertain controls*, Comput. Methods Appl. Mech. Engrg., 213–216 (2012), pp. 152–167.
 - [23] Y. SAAD, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.
 - [24] D. SILVESTER, H. ELMAN, AND A. RAMAGE, *Incompressible Flow and Iterative Solver Software (IFISS) version 3.5*, September 2016.
 - [25] B. SOUSEDÍK AND R. G. GHANEM, *Truncated hierarchical preconditioning for the stochastic Galerkin FEM*, Int. J. Uncertain. Quantif., 4 (2014), pp. 333–348.
 - [26] B. SOUSEDÍK, R. G. GHANEM, AND E. T. PHIPPS, *Hierarchical Schur complement preconditioner for the stochastic Galerkin finite element methods*, Numer. Linear Algebra Appl., 21 (2014), pp. 136–151.
 - [27] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, American Mathematical Society, Providence, Rhode Island, 2010.
 - [28] A. WATHEN AND T. REES, *Chebyshev semi-iteration in preconditioning for problems including the mass matrix*, Electron. Trans. Numer. Anal., 34 (2008-2009), pp. 125–135.
 - [29] D. XIU, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, 2010.
 - [30] R. YOUNGNOI, *Improving convergence in stochastic Galerkin finite element methods via truncation preconditioners*, SIAM J. Sci. Comput., 47 (2025), pp. A1937–A1963.

Appendix A. Auxiliary Results.

A.1. Proof of Lemma 4.2.

Proof. Consider the Rayleigh quotient:

$$\frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T B \mathbf{v}} = \frac{\mathbf{v}^T (B + E) \mathbf{v}}{\mathbf{v}^T B \mathbf{v}} = 1 + \frac{\mathbf{v}^T E \mathbf{v}}{\mathbf{v}^T B \mathbf{v}}.$$

Applying the assumption $|\frac{\mathbf{v}^T E \mathbf{v}}{\mathbf{v}^T B \mathbf{v}}| \leq \delta$ directly yields the bounds. \square

A.2. Proof of Lemma 4.3.

Proof. Let (λ, \mathbf{x}) be an eigenpair satisfying the generalized eigenvalue problem $C\mathbf{x} = \lambda D\mathbf{x}$, with eigenvector $\mathbf{x} \neq \mathbf{0}$. We perform a change of variables by setting $\mathbf{x} = Q\mathbf{y}$. Since Q is nonsingular, $\mathbf{x} \neq \mathbf{0}$ implies that the transformed vector $\mathbf{y} \neq \mathbf{0}$. Substituting $\mathbf{x} = Q\mathbf{y}$ into the original problem gives:

$$C(Q\mathbf{y}) = \lambda D(Q\mathbf{y}).$$

Multiplying from the left by Q^T , we obtain:

$$(Q^T C Q) \mathbf{y} = \lambda (Q^T D Q) \mathbf{y}.$$

This final expression is the generalized eigenvalue problem for the pair $(Q^T C Q, Q^T D Q)$, which is satisfied by the same eigenvalue λ with the transformed eigenvector \mathbf{y} . Therefore, the sets of eigenvalues for both pairs are identical. \square

A.3. Proof of Lemma 4.4.

Proof. Let the SVD of A be $A = U\Sigma V^*$, where U, V are unitary and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ with $\sigma_1 \geq \dots \geq \sigma_n = \sigma_{\min}(A) \geq 0$.

For any nonzero vector \mathbf{v} , let $\mathbf{w} = V^* \mathbf{v}$. Since V is unitary, $\|\mathbf{w}\|_2 = \|\mathbf{v}\|_2$. Then

$$\|A\mathbf{v}\|_2^2 = \|U\Sigma V^* \mathbf{v}\|_2^2 = \|\Sigma \mathbf{w}\|_2^2 = \sum_{i=1}^n \sigma_i^2 |w_i|^2 \geq \sigma_{\min}^2 \sum_{i=1}^n |w_i|^2 = \sigma_{\min}^2 \|\mathbf{w}\|_2^2 = \sigma_{\min}^2 \|\mathbf{v}\|_2^2.$$

Taking square roots on both sides gives $\|A\mathbf{v}\|_2 \geq \sigma_{\min}(A) \|\mathbf{v}\|_2$. \square

A.4. Proof of Lemma 4.5.

Proof. From the triangle inequality, for any vector x , with $\|x\|_2 = 1$, we have

$$\|(A + B)x\|_2 = \|Bx - (-A)x\|_2 \geq \|Bx\|_2 - \|Ax\|_2.$$

Taking the minimum over all unit vectors x on both sides of the inequality, we get

$$\min_{\|x\|_2=1} \|(A + B)x\|_2 \geq \min_{\|x\|_2=1} (\|Bx\|_2 - \|Ax\|_2).$$

Using the property that $\min(f - g) \geq \min(f) - \max(g)$, we obtain

$$\min_{\|x\|_2=1} (\|Bx\|_2 - \|Ax\|_2) \geq \min_{\|x\|_2=1} \|Bx\|_2 - \max_{\|x\|_2=1} \|Ax\|_2.$$

By the definitions of the minimum singular value and the operator norm, the above expression is equivalent to $\sigma_{\min}(A + B) \geq \sigma_{\min}(B) - \|A\|_2$. \square