Random sketching of operators and application to learning preconditioners for model order reduction.

[2] Oleg BALABANOV, [1] Anthony NOUY, [1] Alexandre PASCO

[1]École Centrale Nantes, Nantes Université, France.

[2]International Computer Science Institute; Lawrence Berkeley National Laboratory; Department of Statistics, University of California, Berkeley.

March 2026

Setting

- Hilbert space U with $\dim(U) = n \gg 1$ and $\langle \mathbf{x}, \mathbf{y} \rangle_U = \mathbf{x}^T \mathbf{R}_U \mathbf{y}$.
- Parametric linear operator $\mathbf{A}(\xi) : U \to U'$ and form $\mathbf{b}(\xi) \in U'$.
  – $\mathbf{A}(\xi)$ is ill-conditioned.
- Ultimate goal: approximate $\mathbf{u}(\xi) \in U$, for many $\xi \in \mathcal{P}$, solution to

  $$\mathbf{A}(\xi)\mathbf{u}(\xi) = \mathbf{b}(\xi),$$

  using linear model order reduction (MOR).
  – Deteriorated Galerkin projection and error estimator.
- Intermediate goal: construct $\mathbf{P}(\xi) \in \mathrm{span}\{\mathbf{P}_1, \cdots, \mathbf{P}_p\}$ a linear approximation of $\mathbf{A}(\xi)^{-1}$, tailored to MOR.
  – $\mathbf{P}_i$ are given "implicitly", optimization is challenging.

Classical Galerkin MOR

- Reduced space U_r spanned by orthonormal columns of $\mathbf{U}_r \in \mathbb{R}^{n \times r}$.
- $\hat{\mathbf{u}} = \mathbf{U}_r \hat{\mathbf{a}}$ with $\hat{\mathbf{a}} \in \mathbb{R}^r$ solution to the $r \times r$ linear system,

$$\left(\mathbf{U}_r^T \mathbf{A} \mathbf{U}_r\right) \hat{\mathbf{a}} = \mathbf{U}_r^T \mathbf{b}.$$

- Quasi-optimality from Cea's Lemma,

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_U \leq \frac{\sigma_1(\mathbf{A})}{\alpha(\mathbf{A})} \|\mathbf{u} - \mathbf{\Pi}_{U_r} \mathbf{u}\|_U.$$

- Accuracy of the error estimator $\|\mathbf{A}\hat{\mathbf{u}} - \mathbf{b}\|_{U'}$,

$$\sigma_n(\mathbf{A}) \leq \frac{\|\mathbf{A}\hat{\mathbf{u}} - \mathbf{b}\|_{U'}}{\|\mathbf{u} - \hat{\mathbf{u}}\|_U} \leq \sigma_1(\mathbf{A}).$$

– Ill conditioning deteriorates quasi optimality and error estimation.

Preconditioned Galerkin MOR

- Reduced space U_r spanned by orthonormal columns of $\mathbf{U}_r \in \mathbb{R}^{n\times r}$.
- $\hat{\mathbf{u}} = \mathbf{U}_r \hat{\mathbf{a}}$ with $\hat{\mathbf{a}} \in \mathbb{R}^r$ solution to the $r \times r$ linear system,

$$\left(\mathbf{U}_r^T \mathbf{R}_U \mathbf{P} \mathbf{A} \mathbf{U}_r\right) \hat{\mathbf{a}} = \mathbf{U}_r^T \mathbf{R}_U \mathbf{P} \mathbf{b}.$$

- Quasi-optimality from Cea's Lemma,

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_U \leq \frac{\sigma_1(\mathbf{P}\mathbf{A})}{\alpha(\mathbf{P}\mathbf{A})} \|\mathbf{u} - \mathbf{\Pi}_{U_r}\mathbf{u}\|_U.$$

- Accuracy of the error estimator $\|\mathbf{P}\mathbf{A}\hat{\mathbf{u}} - \mathbf{P}\mathbf{b}\|_U$,

$$\sigma_n(\mathbf{P}\mathbf{A}) \leq \frac{\|\mathbf{P}\mathbf{A}\hat{\mathbf{u}} - \mathbf{P}\mathbf{b}\|_{U'}}{\|\mathbf{u} - \hat{\mathbf{u}}\|_U} \leq \sigma_1(\mathbf{P}\mathbf{A}).$$

General purpose in operator norm

- General purpose: discrepancy in operator norm,

$$\|\mathbf{I} - \mathbf{PA}\|_{U,U} := \sup_{\mathbf{x} \in U,\ \|\mathbf{x}\|_U = 1} \|(\mathbf{I} - \mathbf{PA})\mathbf{x}\|_U$$

- Bound on the inf-sup constants of $\mathbf{PA}$,

$$1 - \|\mathbf{I} - \mathbf{PA}\|_{U,U} \leq \sigma_n(\mathbf{PA}) \leq \sigma_1(\mathbf{PA}) \leq 1 + \|\mathbf{I} - \mathbf{PA}\|_{U,U}.$$

- Problem: requires $\|\mathbf{I} - \mathbf{PA}\|_{U,U} < 1$, but online computation and optimization of $\mathbf{P} \mapsto \|\mathbf{I} - \mathbf{PA}\|_{U,U}$ is untractable.

General purpose in HS norm

- Alternative: Hilbert-Schmidt norm, given o.n.b $(\mathbf{x}_i)_{1\leq i\leq n}$ of U,

$$\|\mathbf{I}-\mathbf{PA}\|_{HS(U,U)}^2 := \sum_{i=1}^{n} \|(\mathbf{I}-\mathbf{PA})\mathbf{x}_i\|_U^2 \geq \|\mathbf{I}-\mathbf{PA}\|_{U,U}^2.$$

- Minimization of a linear least-squares problem,

$$\min_{\mathbf{P}\in\text{span}\{\mathbf{P}_1,\cdots,\mathbf{P}_p\}} \|\mathbf{I}-\mathbf{PA}\|_{HS(U,U)}^2.$$

- Problem 1: evaluating $\|\mathbf{I}-\mathbf{PA}\|_{HS(U,U)}$ is (very) costly.
  $\rightarrow$ [Zahm and Nouy, 2016] used random estimator.

- Problem 2: HS norm can be much larger than operator norm,

$$\frac{1}{\sqrt{\dim(U)}}\|\cdot\|_{HS(U,U)} \leq \|\cdot\|_{U,U} \leq \|\cdot\|_{HS(U,U)}.$$

$\rightarrow$ Seminorms tailored to MOR.

MOR purpose operator seminorm: preconditioned Galerkin

- MOR purpose: discrepancy in operator seminorms,

$$\|\mathbf{I} - \mathbf{PA}\|_{U,U_r} := \|\mathbf{\Pi}_{U_r}(\mathbf{I} - \mathbf{PA})\|_{U,U}.$$

- Computable with $(\mathbf{v}_i)_{1 \leq i \leq r}$ o.n.b of U_r,

$$\|\mathbf{I} - \mathbf{PA}\|_{U,U_r}^2 = \sigma_1 \left(\mathbf{U}_r^T \mathbf{R}_U (\mathbf{I} - \mathbf{PA}) \mathbf{R}_U (\mathbf{I} - \mathbf{PA})^T \mathbf{R}_U \mathbf{U}_r \right).$$

+ Evaluating $\|\mathbf{I} - \mathbf{PA}\|_{U,U_r}^2$ is not very costly.
− Sensitive to round-off errors due to $(\mathbf{I} - \mathbf{PA})(\mathbf{I} - \mathbf{PA})^T$.
− Minimizing $\mathbf{P} \mapsto \|\mathbf{I} - \mathbf{PA}\|_{U,U_r}^2$ is not a linear l-s problem.

Proposition

The preconditioned Galerkin projection $\hat{\mathbf{u}}$ on U_r satisfies

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_U \leq \frac{1}{1 - \|\mathbf{I} - \mathbf{PA}\|_{U,U_r}} \|\mathbf{u} - \mathbf{\Pi}_{U_r} \mathbf{u}\|_U.$$

MOR purpose operator seminorm: error estimator

- Assume given $U_m \supset U_r$, $r \leq m \ll n$, and for some $\tau \in (0,1)$,

$$\|\mathbf{u} - \mathbf{\Pi}_{U_m}\mathbf{u}\|_U \leq \tau \|\mathbf{u} - \hat{\mathbf{u}}\|_U.$$

Proposition

Under the above assumption, $\|\mathbf{\Pi}_{U_m}\mathbf{P}(\mathbf{A}\hat{\mathbf{u}} - \mathbf{b})\|_U$ *satisfies,*

$$\frac{\|\mathbf{\Pi}_{U_m}\mathbf{P}(\mathbf{A}\hat{\mathbf{u}} - \mathbf{b})\|_U}{1 + (1+\tau)\|\mathbf{I} - \mathbf{PA}\|_{U,U_m}} \leq \|\mathbf{u} - \hat{\mathbf{u}}\|_U \leq \frac{\|\mathbf{\Pi}_{U_m}\mathbf{P}(\mathbf{A}\hat{\mathbf{u}} - \mathbf{b})\|_U}{\sqrt{1-\tau^2} - (1+\tau)\|\mathbf{I} - \mathbf{PA}\|_{U,U_m}}.$$

- $U_r \subset U_m \subset U$ thus $\|\cdot\|_{U,U_r} \leq \|\cdot\|_{U,U_m} \leq \|\cdot\|_{U,U}$.

MOR purpose HS seminorm

- Alternative: Hilbert-Schmidt seminorm,

$$\|\mathbf{I} - \mathbf{PA}\|_{HS(U,U_m)} := \|\mathbf{\Pi}_{U_m}(\mathbf{I} - \mathbf{PA})\|_{HS(U,U)},$$

- HS seminorms are almost equivalent to operator seminorms,

$$\frac{1}{\sqrt{m}}\|\cdot\|_{HS(U,U_m)} \leq \|\cdot\|_{U,U_m} \leq \|\cdot\|_{HS(U,U_m)}.$$

- Computable with $\mathbf{U}_m \in \mathbb{R}^{n\times m}$ with columns forming an o.n.b of U_m,

$$\|\mathbf{I} - \mathbf{PA}\|^2_{HS(U,U_m)} = \mathrm{Tr}\left(\mathbf{U}_m^T\mathbf{R}_U(\mathbf{I} - \mathbf{PA})(\mathbf{I} - \mathbf{PA})^T\mathbf{R}_U\mathbf{U}_m\right),$$

+ Evaluating $\|\mathbf{I} - \mathbf{PA}\|^2_{HS(U,U_m)}$ is not very costly.
+ Minimizing $\mathbf{P} \mapsto \|\mathbf{I} - \mathbf{PA}\|^2_{HS(U,U_m)}$ is a linear l-s problem.

− Sensitive to round-off errors due to $(\mathbf{I} - \mathbf{PA})(\mathbf{I} - \mathbf{PA})^T$.

Quality measures: take home messages

- Construct $\mathbf{P} \simeq \mathbf{A}^{-1}$ by minimizing $\|\mathbf{I} - \mathbf{PA}\|$.
- Operator norms yields sharper theoretical results.
- HS norms yields a linear least-squares problem.
- Tailoring to the reduced space yields good approximation of operator seminorms by HS seminorms.

Random sketching for vectors

- $\boldsymbol{\Theta} \in \mathbb{R}^{k \times n}$ is an ε-**subspace embedding** for a subspace $V \subset \mathbb{R}^n$ if

$$\forall \mathbf{v} \in V, \; \left| \|\boldsymbol{\Theta}\mathbf{v}\|_2^2 - \|\mathbf{v}\|_U^2 \right| \leq \varepsilon \|\mathbf{v}\|^2.$$

  $\rightarrow$ [Woodruff, 2014, Martinsson and Tropp, 2020, Murray et al., 2023].

- For $\boldsymbol{\Theta}$ with $k = \mathcal{O}(\varepsilon^{-2}(\dim(V) + \log(\delta^{-1})))$ rows as independent Gaussian vectors with covariance depending on U,

$$\mathbb{P}\left[\forall \mathbf{v} \in V, \; \left| \|\boldsymbol{\Theta}\mathbf{v}\|_2^2 - \|\mathbf{v}\|_U^2 \right| \leq \varepsilon \|\mathbf{v}\|^2\right] \geq 1 - \delta.$$

  We say that $\boldsymbol{\Theta}$ is an **oblivious subspace embedding**.
  $\rightarrow$ In practice, use structured or sparse embedding for efficiency, so that computing $\boldsymbol{\Theta}\mathbf{v}$ costs $n \log(n)$.

- Problem: $\mathbf{P}_i$ are **inverses** of (sparse) matrices, thus only given "implicitely", via application to a vector.
  $\rightarrow$ Cannot use classical embeddings for vectors.

Random sketching for HS operators

For $\mathbf{\Omega}$, $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$ with k row as "classical" embeddings for vectors, then for an HS operator $\mathbf{X}$ we consider,

$$\mathbf{\Theta}(\mathbf{X}) := \mathbf{\Gamma}\mathrm{vec}(\mathbf{\Omega}\mathbf{X}\mathbf{\Sigma}^T) \in \mathbb{R}^k.$$

→ Computed with k operator-vector application.

→ Adaptable for seminorms.

Proposition

Let $\mathcal{X}$ a subspace of HS operators. Let $\mathbf{\Omega}$, $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$ random matrices with $k = \mathcal{O}(\varepsilon^{-2}(\dim(\mathcal{X}) + \log(\delta^{-1})))$ rows as independent Gaussian vectors with suitable covariance. Then, for any d-dimensional subspace $\mathcal{X}$ of HS operators,

$$\mathbb{P}\left[\forall \mathbf{X} \in \mathcal{X}, \left| \|\mathbf{\Theta}(\mathbf{X})\|_2^2 - \|\mathbf{X}\|_{HS}^2 \right| \leq \varepsilon \|\mathbf{X}\|_{HS}^2 \right] \geq 1 - \delta,$$

→ In practice, use $\mathbf{\Sigma}$ as Gaussian, and $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ as structured or sparse.

Sketched measures of quality

- The discrepancy matrix $\mathbf{I} - \mathbf{PA}$ lies in a $(p+1)$-dimensional space,

$$\mathbf{I} - \mathbf{PA} \in \mathrm{span}\{\mathbf{I}, \mathbf{P}_1\mathbf{A}, \cdots, \mathbf{P}_p\mathbf{A}\}.$$

$\rightarrow$ Small sketch size $k = \mathcal{O}(\varepsilon^{-2}(p + \log(\delta^{-1})))$.

- Minimizing discrepancy in sketched HS (semi)norms is a **small** linear least-squares problem,

$$\min_{\mathbf{P} \in \mathrm{span}\{\mathbf{P}_1, \cdots, \mathbf{P}_p\}} \|\mathbf{\Theta}(\mathbf{I} - \mathbf{PA})\|_2^2 = \min_{\mathbf{a} \in \mathbb{R}^p} \|\mathbf{h} - \mathbf{Wa}\|_2^2,$$

$$\mathbf{h} := \mathbf{\Theta}(\mathbf{I}) \in \mathbb{R}^k \quad \text{and} \quad \mathbf{W} := (\mathbf{\Theta}(\mathbf{P}_1\mathbf{A}), \cdots, \mathbf{\Theta}(\mathbf{P}_p\mathbf{A})) \in \mathbb{R}^{k \times p}.$$

$\rightarrow$ Solve with stable method (SVD, QR) instead of normal equation.
$\rightarrow$ Near optimality with probability $1 - \delta$.

Practical aspects

- Assume **parameter separability**, $\mathbf{A}(\xi) = \sum_{j=1}^{q} \theta_j(\xi)\mathbf{A}_j$.
  $\rightarrow$ Problem formulation, Empirical Interpolation [Maday et al., 2009].
- The columns $\mathbf{\Theta}(\mathbf{P}_i\mathbf{A}(\xi))$ of $\mathbf{W}(\xi)$ are also parameter separable,

$$\mathbf{\Theta}(\mathbf{P}_i\mathbf{A}(\xi)) = \sum_{j=1}^{q} \theta_j(\xi)\mathbf{\Theta}(\mathbf{P}_i\mathbf{A}_j) \in \mathbb{R}^k$$

  $\rightarrow$ Compute efficiently $\mathbf{\Theta}(\mathbf{P}_i\mathbf{A}_j)$ independently of ξ.
- If $\mathbf{b}$ is parameter separable, then preconditioned Galerkin system and preconditioned error estimator are also **parameter separable**.
- Set $\mathbf{P}_i = \mathbf{A}(\xi_i)^{-1}$ with $\xi_i \in \mathcal{P}$ selected with a greedy algorithm.

Random Sketching for operators: take home messages

- Sketch an operator from a small number k of input-output queries.
- Solve sketched linear least-squares in HS norms with stable methods at small cost.
- Near-optimality with high user-defined probability and accuracy.

Numerical example

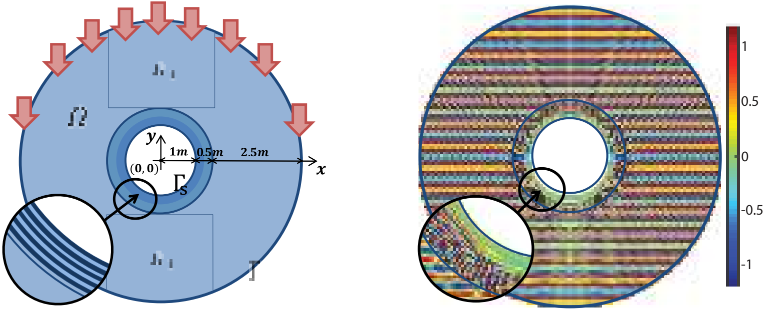

Figure: [Balabanov and Nouy, 2021] Wave scattering in 2D with a perfect scatterer covered in an invisibility cloak composed of layers of homogeneous isotropic materials, with $n \geq 400\ 000$ dofs. Left: Geometry of the problem. Right: real part of random snapshot. Sketch sizes $k = 1024$.

Numerical example: absolute error

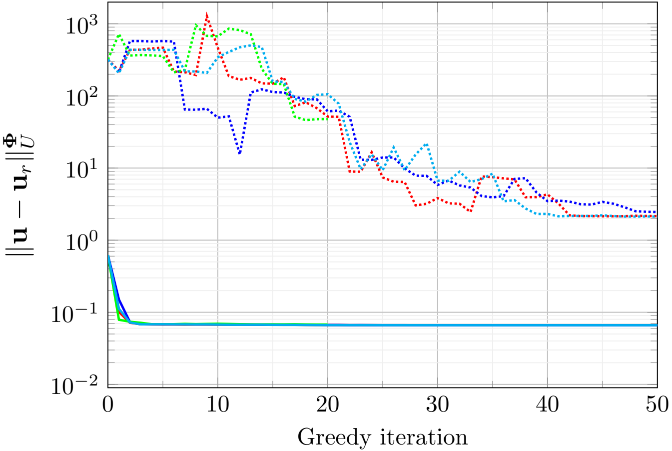

Figure: Quantiles on test sets (90% continuous, 100% dotted) for the preconditioned Galerkin projection, along the greedy construction of the preconditioner space. Three sketched greedy criterions: green is $HS(U, U)$, red is $HS(U_m, U_m)$, blue is $HS(U, U_m)$, cyan is weighted sum.

Numerical example: Galerkin quasi optimality

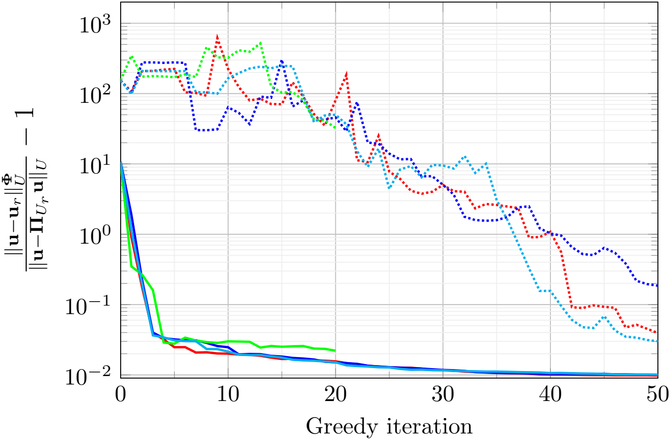

Figure: Quantiles on test sets (90% continuous, 100% dotted) for the preconditioned Galerkin projection, along the greedy construction of the preconditioner space. Three sketched greedy criterions: green is $HS(U,U)$, red is $HS(U_m,U_m)$, blue is $HS(U,U_m)$, cyan is weighted sum.

Numerical example: error estimator

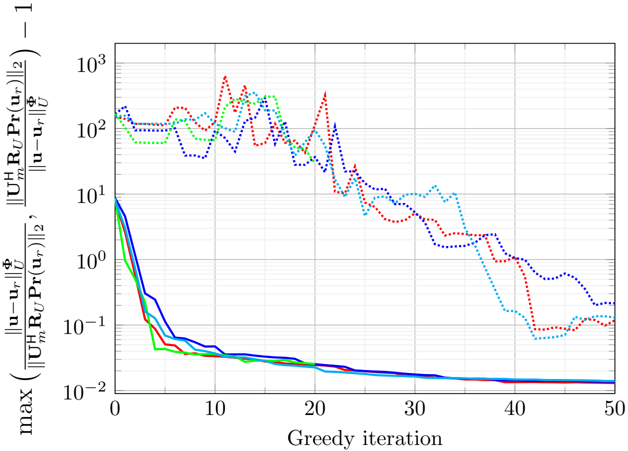

Figure: Quantiles on test sets (90% continuous, 100% dotted) for the preconditioned Galerkin projection, along the greedy construction of the preconditioner space. Three sketched greedy criterions: green is $HS(U, U)$, red is $HS(U_m, U_m)$, blue is $HS(U, U_m)$, cyan is weighted sum.

Big picture: three linear approximations

Classical approaches: linear approximations of...

- The solution, $\mathbf{u}(\xi) \simeq \hat{\mathbf{u}}(\xi) \in \mathrm{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_r\} = U_r$.
  $\rightarrow$ Reduced Basis [Noor and Peters, 1980, Veroy et al., 2003].
- The operator, $\mathbf{A}(\xi) \in \mathrm{span}\{\mathbf{A}_1, \cdots, \mathbf{A}_q\}$.
  $\rightarrow$ Problem formulation, Empirical Interpolation [Maday et al., 2009].

Not classical approach: linear approximation of the inverse operator,

$$\mathbf{A}(\xi)^{-1} \simeq \mathbf{P}(\xi) \in \mathrm{span}\{\mathbf{P}_1, \cdots, \mathbf{P}_p\}, \qquad \min_{\mathbf{P} \in \mathrm{span}\{\mathbf{P}_1, \cdots, \mathbf{P}_p\}} \|\mathbf{I} - \mathbf{P}\mathbf{A}(\xi)\|,$$

with for example $\mathbf{P}_i = \mathbf{A}(\xi_i)^{-1}$.
$\rightarrow$ General purpose with randomized methods [Zahm and Nouy, 2016].
$\rightarrow$ MOR purpose for sharper bounds.
$\rightarrow$ Random sketching for efficient and numerically stable computations.

Perspectives

Perspectives for MOR:

- Greedy algorithm constructing $(\mathbf{P}_i)_{1\leq i\leq p}$ and U_r at the same time, as in [Zahm and Nouy, 2016].
- Nonlinear construction of $\mathbf{P}(\xi)$ (e.g., piecewise linear, dictionary, ...).

Perspectives for random sketching:

- Sketching of operators for other settings (e.g., eigenvalue problems, domain decomposition, sketching low-rank matrices and tensors, ...).

Thank you for your attention.

References I

Balabanov, O. and Nouy, A. (2021).
Randomized linear algebra for model reduction—part II: Minimal residual methods and dictionary-based approximation.
Adv Comput Math, 47(2):26.

Maday, Y., Cuong Nguyen, N., T. Patera, A., and H. Pau, S. (2009).
A general multipurpose interpolation procedure: The magic points.
Communications on Pure & Applied Analysis, 8(1):383–404.

Martinsson, P.-G. and Tropp, J. A. (2020).
Randomized numerical linear algebra: Foundations and algorithms.
Acta Numerica, 29:403–572.

Murray, R., Demmel, J., Mahoney, M. W., Erichson, N. B., Melnichenko, M., Malik, O. A., Grigori, L., Luszczek, P., Dereziński, M., Lopes, M. E., Liang, T., Luo, H., and Dongarra, J. (2023).
Randomized Numerical Linear Algebra : A Perspective on the Field With an Eye to Software.

Noor, A. K. and Peters, J. M. (1980).
Reduced Basis Technique for Nonlinear Analysis of Structures.
AIAA Journal, 18(4):455–462.

Veroy, K., Prud'homme, C., Rovas, D., and Patera, A. (2003).
A Posteriori Error Bounds for Reduced-Basis Approximation of Parametrized Noncoercive and Nonlinear Elliptic Partial Differential Equations.
In *16th AIAA Computational Fluid Dynamics Conference*, Orlando, Florida. American Institute of Aeronautics and Astronautics.

References III

Woodruff, D. P. (2014).
Sketching as a Tool for Numerical Linear Algebra.
FNT in Theoretical Computer Science, 10(1-2):1–157.

Zahm, O. and Nouy, A. (2016).
Interpolation of Inverse Operators for Preconditioning Parameter-Dependent Equations.
SIAM J. Sci. Comput., 38(2):A1044–A1074.