# A Brief Introduction to Reduced Order Models

Max Hirsch*

Week of March 10th, 2025

## 1 Goals

- Learn how to simulate PDEs faster using the proper orthogonal decomposition and the discrete empirical interpolation method.
- Learn the importance of linear algebra for developing methods for solving PDEs and learn fundamental proofs in low rank approximation theory.

## 2 Allen-Cahn Equation

Consider the 1-D Allen-Cahn equation

$$\begin{cases} u_t - \Delta u = f(u), & (t,x) \in (0,T] \times \Omega \\ u(0,x) = u_0(x;\mu), & x \in \overline{\Omega} \\ u(t,0) = u(t,1) = 0, & t \in [0,T], \end{cases} \tag{1}$$

where $\Omega = [0,1]$ is our spatial domain, $f(u) = u - u^3$, and $u_0$ is an initial condition parameterized by some parameter $\mu \in \mathcal{P}$. Note that $f(u) = -F'(u)$, where $F(u) = (u^2-1)^2/4$. See Figure 1 for a visualization of $F$.

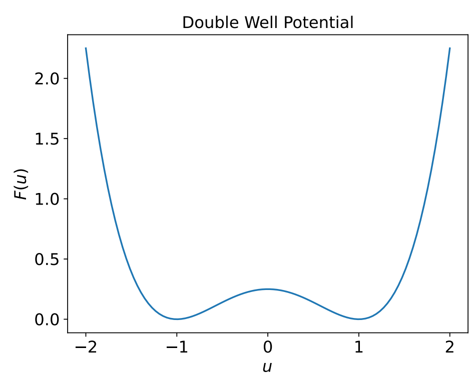


Figure 1: Plot of the double well potential $F(u)$.

*Much of these notes is an adaptation of notes from Daniel Kressner's course on Low Rank Approximation Techniques.

# 3 Discretization

To discretize this equation, consider a number of time steps $N_t$ and let $\Delta t = T/N_t$. Then define $t_n = n\Delta t$ so that we have $[0,T]$ discretized by $0 = t_0 < t_1 < \cdots < t_{N_t} = T$. Similarly discretize $\Omega$ with $\Delta x = 1/N_x$ so that $x_n = n\Delta x$ and $0 = x_0 < x_1 < \cdots < x_{N_s} = 1$.

Now for simplicity, consider the following convex splitting scheme for solving the Allen-Cahn equation (1):

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} - \frac{u_{j+1}^{n+1/2} - 2u_j^{n+1/2} + u_{j-1}^{n+1/2}}{\Delta x^2} = u_j^n - (u_j^{n+1})^3, \quad n = 1, 2, \ldots, N_t, \quad j = 1, \ldots, N_x - 1, \tag{2}$$

with $a^{n+1/2} = (a^n + a^{n+1})/2$, and

$$u_0^n = u_{N_x}^n = 0 \quad \forall n = 0, 1, 2, \ldots, N_t, \quad u_j^0 = u_0(x_j; \mu) \quad \forall j = 0, 1, \ldots, N_x.$$

Then we can write the solution to (2) and the initial and boundary conditions together as

$$\frac{u_h^{n+1} - u_h^n}{\Delta t} - Au_h^{n+1/2} = u_h^n - (u_h^{n+1})^3,$$

where the cube is understood componentwise and

$$A = \begin{bmatrix} 1 & 0_{1\times(N_x-2)} & 0 \\ 0_{(N_x-2)\times 1} & B & 0_{(N_x-2)\times 1} \\ 0 & 0_{1\times(N_x-2)} & 1 \end{bmatrix}, \quad B = \frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & & \ddots & \\ & & & 1 & -2 \end{bmatrix}.$$

Example solutions are shown in Figure 2.

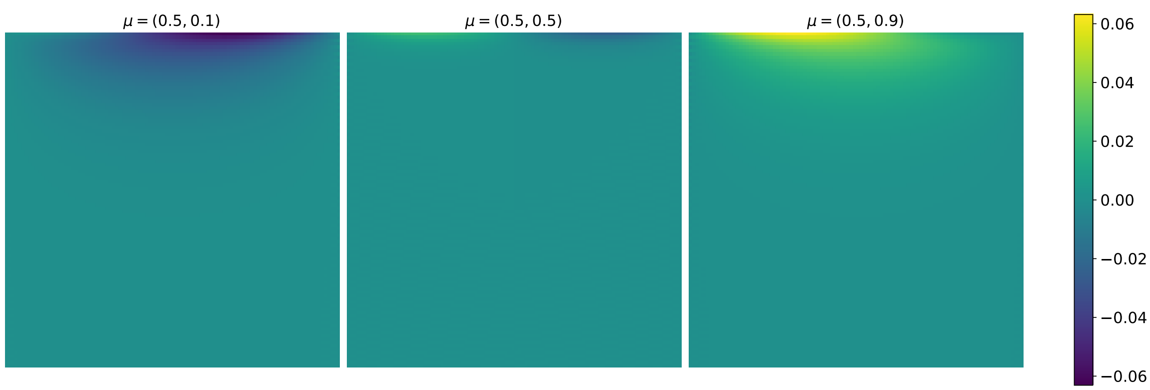


Figure 2: Solutions to the discretized Allen-Cahn equation (2) with initial condition $u_0(x;\mu) = \mu_1 x(x-1)(x-\mu_2)$.

In general, solving a nonlinear PDE with a fine discretization can be very slow, and we may want solutions for many values of the parameter $\mu$, so we will have to simulate the equation repeatedly many times. This motivates us to look for ways to approximate our discretization with something which will be faster to solve.

# 4 Proper Orthogonal Decomposition

## 4.1 General Low Rank Approximation Theory

**Definition 1** (Frobenius norm). *The Frobenius norm of an $m \times n$ matrix $A$, denoted $\|A\|_F$, is defined by*

$$\|A\|_F = \sqrt{\operatorname{tr}(A^\top A)} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}.$$

**Exercise 1.** *Check that $\langle A, B \rangle = \operatorname{tr}(A^\top B)$ defines an inner product on the space of $m \times n$ matrices. In particular, the Frobenius norm is the norm induced by this inner product.*

**Definition 2** (Rank $k$ Truncation). *Given $A \in \mathbb{R}^{m \times n}$ with SVD $A = U \Sigma V^\top$, its rank $k$ truncation $\mathcal{T}_k(A)$ is defined by*

$$\mathcal{T}_k(A) = U \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix} V^\top = U_k \Sigma_k V_k^\top,$$

*where $\Sigma_k$ is the diagonal matrix $\operatorname{diag}(\sigma_1(A), \sigma_2(A), \ldots, \sigma_k(A))$, $U_k$ is the first $k$ columns of $U$, and $V_k$ is the first $k$ columns of $V$. In other words, we replace all but the first $k$ singular values in $\Sigma$ with zeros.*

**Definition 3** (Spectral Norm). *The spectral norm of an $m \times n$ matrix $A$, denoted $\|A\|_2$, is defined by*

$$\|A\|_2 = \sup\{\|Ax\|_2 : \|x\|_2 \leq 1\} = \sigma_{\max}(A),$$

*where $\sigma_{\max}(A)$ is the maximum singular value of $A$.*

**Exercise 2.** *Show that the second equality in the definition above holds, that $\|A\|_F \leq \min(m,n)^{1/2} \|A\|_2$, and that $\|A\|_2 \leq \|A\|_F$. Hint: SVD and unitary invariance of Euclidean, spectral, and Frobenius norms.*

Then we can compute the error of the rank $k$ truncation of $A$ in the Frobenius norm as

$$\|A - \mathcal{T}_k(A)\|_F = \|U \Sigma V^\top - U \mathcal{T}_k(\Sigma) V^\top\|_F = \|\Sigma - \mathcal{T}_k(\Sigma)\|_F = \sqrt{\sum_{i=k+1}^{\min\{m,n\}} \sigma_i(A)^2}$$

and in the spectral norm as

$$\|A - \mathcal{T}_k(A)\|_2 = \|U \Sigma V^\top - U \mathcal{T}_k(\Sigma) V^\top\|_2 = \|\Sigma - \mathcal{T}_k(\Sigma)\|_2 = \sigma_{k+1}(A).$$

**Theorem 1** (Von-Neumann Trace Inequality). *For $m \geq n$, let $A, B \in \mathbb{R}^{m \times n}$ have singular values $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$ and $\sigma_1(B) \geq \cdots \geq \sigma_n(B)$ respectively. Then*

$$|\langle A, B \rangle| \leq \sigma_1(A)\sigma_1(B) + \cdots + \sigma_n(A)\sigma_n(B).$$

*Proof.* We first prove the result for

$$A = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Indeed, partition

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

with $B_{11} \in \mathbb{R}^{k\times k}$. Take the SVD of $B = U\Sigma V^\top$ and write

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} U_{11} & \dots & U_{1m} \\ & * & \end{bmatrix}, \quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} V_{11} & \dots & V_{1n} \\ & * & \end{bmatrix},$$

with $U_1 \in \mathbb{R}^{k\times m}$ and $V_1 \in \mathbb{R}^{k\times n}$. Then

$$\operatorname{tr}(B_{11}) = \operatorname{tr}(U_1\Sigma_1 V_1^\top) = \operatorname{tr}(V_1^\top U_1 \Sigma_1) = \alpha_1\sigma_1(B) + \cdots + \alpha_n\sigma_n(B),$$

where $\alpha_j = U_{1j}^\top V_{1j}$. Note that $|\alpha_j| \le \|U_{1j}\|\|V_{1j}\| \le 1$ by the Cauchy-Schwarz inequality. Also,

$$|\langle U_1, V_1\rangle| \le \|U_1\|_F\|V_1\|_F \le \sqrt{k}\|U_1\|_2 \cdot \sqrt{k}\|V_1\|_2 \le k\|U\|_2\|V\|_2 = k.$$

(Indeed, the last inequality is shown by considering $x^\top U^\top Ux = x^\top U_1^\top U_1 x + x^\top U_2^\top U_2 x \ge x^\top U_1^\top U_1 x$ and similarly for $V$.) It follows that

$$\operatorname{tr}(U_1V_1^\top) = \operatorname{tr}(V_1^\top U_1) = \alpha_1 + \cdots + \alpha_n \le k.$$

Thus,

$$\operatorname{tr}(B_{11}) \le \max_{\substack{\alpha_1,\dots,\alpha_n \le 1: \\ \alpha_1+\cdots+\alpha_n \le k}} \sum_{i=1}^n \alpha_i\sigma_i(B) \le \sigma_1(B) + \cdots + \sigma_k(B),$$

since $\sigma_1(B) \ge \cdots \ge \sigma_n(B)$. This proves the special case. Now consider general $A$.

Take the SVD of $A = U_A\Sigma_A V_A^\top$. Then

$$\langle A, B\rangle = \operatorname{tr}(B^\top U_A\Sigma_A V_A^\top) = \langle \Sigma_A, \underbrace{U_A^\top B V_A}_{\widetilde{B}}\rangle,$$

so without loss of generality,

$$\begin{aligned} A = \Sigma_A &= \begin{bmatrix} \sigma_1(A) & & \\ & \ddots & \\ & & \sigma_n(A) \end{bmatrix} \\ &= \sigma_n(A)\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} + \underbrace{(\sigma_{n-1}(A) - \sigma_n(A))}_{\ge 0} \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{bmatrix} \\ &\quad + \cdots + \underbrace{(\sigma_1(A) - \sigma_2(A))}_{\ge 0} \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \\ &= \sum_{k=1}^n \alpha_k E_k, \end{aligned}$$

where $\alpha_k = \sigma_k(A) - \sigma_{k+1}(A)$ ($\sigma_{n+1}(A) := 0$) and $E_k = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$. We then have

$$\begin{aligned}
\langle A, B \rangle = \left\langle \sum_{k=1}^{n} \alpha_k E_k, B \right\rangle &= \sum_{k=1}^{n} \alpha_k \langle E_k, B \rangle \\
&\leq \sum_{k=1}^{n} (\sigma_k(A) - \sigma_{k+1}(A))(\sigma_1(B) + \cdots + \sigma_k(B)) \\
&= \sigma_1(A)\sigma_1(B) + \cdots + \sigma_n(A)\sigma_n(B),
\end{aligned}$$

as desired.

□

As a consequence of the Von-Neumann trace inequality, we have

$$\begin{aligned}
\|A - B\|_F^2 &= \langle A - B, A - B \rangle \\
&= \|A\|_F^2 - 2\langle A, B \rangle + \|B\|_F^2 \\
&\geq \|A\|_F^2 - 2\sum_{i=1}^{n} \sigma_i(A)\sigma_i(B) + \|B\|_F^2 \\
&= \sum_{i=1}^{n} (\sigma_i(A) - \sigma_i(B))^2.
\end{aligned}$$

**Theorem 2** (Schmidt-Mirsky)**.** *Let $A \in \mathbb{R}^{m \times n}$. Then*

$$\|A - \mathcal{T}_k(A)\| = \min\{\|A - B\| : B \in \mathbb{R}^{m \times n} \text{ has rank at most } k\}$$

*holds for any unitarily invariant norm $\|\cdot\|$.*

*Proof.* (For $\|\cdot\|_F$). We apply the consequence of the Von-Neumann trace inequality. Indeed, for any rank $k$ matrix $B$, we have that $\sigma_i(B) = 0$ for $i \geq k+1$, so

$$\|A - B\|_F^2 = \sum_{i=1}^{k} (\sigma_i(A) - \sigma_i(B))^2 + \sum_{i=k+1}^{n} \sigma_i(A)^2 \geq \sum_{i=k+1}^{n} \sigma_i(A)^2 = \|A - \mathcal{T}_k(A)\|_F^2.$$

□

Now say we want to find a matrix $Q \in \mathbb{R}^{m \times k}$ with orthonormal columns such that

$$\mathrm{range}(Q) = \mathrm{range}(A).$$

Then $QQ^\top$ is the projection onto $\mathrm{range}(Q)$ and $I - QQ^\top$ is the projection onto $\mathrm{range}(Q)^\perp$, i.e. the projection error. So we want to minimize the projection error

$$\|(I - QQ^\top)A\| = \|A - QQ^\top A\|$$

for $\|\cdot\|$ unitarily invariant, say the Frobenius or spectral norm. Now because $\operatorname{rank}(QQ^\top) \le k$, we have that $\operatorname{rank}(QQ^\top A) \le k$, so by Schmidt-Mirsky,

$$\|A - QQ^\top A\| \ge \|A - \mathcal{T}_k(A)\|.$$

If we take $Q = U_k$ the $k$ left singular vectors of $A$, we obtain

$$U_k U_k^\top A = U_k U_k^\top U \Sigma V^\top = U_k \Sigma_k V_k^\top = \mathcal{T}_k(A),$$

so we see that $Q = U_k$ is optimal.

## 4.2 Proper Orthogonal Decomposition

Returning to our discretized PDE, we want to represent our solution in a lower dimensional space in which it might be faster to solve our PDE. The idea is as follows:

1. Simulate the equation (1) for multiple parameters $\mu$ using the discretization (2) and collect the data into the "snapshot" matrix:
$$S = \begin{bmatrix} u_h^0(\mu_1) & \dots & u_h^{N_t}(\mu_1) & \dots & u_h^0(\mu_p) & \dots & u_h^{N_t}(\mu_p) \end{bmatrix}.$$
2. Compute SVD of $S = U\Sigma V^\top$ and truncate to the $r$th principal subspace of $S$ with $r \ll N_x$ (i.e., retain the first $r$ columns $U_r$ of $U$).
3. Assume that $u_h^n \approx U_r U_r^\top u_h^n = U_r \widetilde{u}^n$ and project the discretized PDE onto $\operatorname{range}(U_r)$:
$$\frac{\widetilde{u}^{n+1} - \widetilde{u}^n}{\Delta t} - U_r^\top A U_r \widetilde{u}^{n+1/2} = U_r^\top (U_r \widetilde{u}^n - (U_r \widetilde{u}^{n+1})^3). \tag{3}$$

We've already seen that this projection onto $\operatorname{range}(U_r)$ is the best rank $r$ linear projection onto the column space of the snapshot matrix. Now note that $U_r^\top A U_r$ is an $r \times r$ matrix which is much smaller than $A$, so we hope that this equation is faster to solve–the cost of evaluating $U_r^\top A U_r \widetilde{u}^n$ is $\mathcal{O}(r^2)$ since $U_r^\top A U_r$ can be precomputed and saved, while the cost of $Au_h^n$ is $\mathcal{O}(N_x^2)$. However, the computational complexity of evaluating the nonlinear term on the right hand side still depends on the large dimension $N_x$ since it cannot be precomputed like the linear term. We will now look at one method for speeding up the evaluation of the nonlinear term.

# 5 Discrete Empirical Interpolation

First, we give the main idea of discrete empirical interpolation. Let $g$ be some componentwise nonlinearity, in an abuse of notation,

$$g(u_1, u_2, \dots, u_k) = (g(u_1), g(u_2), \dots, g(u_k)).$$

Here we mean that $g$ acts on components of the vector individually, as in our polynomial nonlinearity in the Allen-Cahn equation. We want an approximation of $g$ that is faster to compute. Maybe we can select rows to compute and others we skip computing (maybe they are somehow able to be approximated from the other components). Let $I = \{i_1, \ldots, i_k\}$ be some selected component indices for which we want to compute $g$ and let

$$\mathbb{S} = \begin{bmatrix} e_{i_1} & \ldots & e_{i_k} \end{bmatrix},$$

where $e_i$ denotes the $i$th standard basis vector in $\mathbb{R}^{N_x}$. Then $\mathbb{S}^\top g$ is a $k$ dimensional vector with components equal to the selected components of $g$. For example, if $N_x = 3$ and $k = 2$ then

$$\mathbb{S}^\top g = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}^\top \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_3 \end{bmatrix},$$

that is, this $\mathbb{S}$ selects the first and third components of $g$. The key observation is that if we take some suitable projection of $g$ containing $\mathbb{S}^\top$, then we only need to compute the selected indices of $g$. Note that if $V \in \mathbb{R}^{N_x \times k}$ is some orthonormal basis, then

$$V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top$$

is an oblique projection onto the column space of $V$ (check $(V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top)(V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top) = V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top$). Furthermore, we have that

$$\mathbb{S}^\top(I - V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top) = 0,$$

so $\|\mathbb{S}^\top(g - V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top g)\| = 0$. This means that $g$ is "interpolated" exactly at the selected indices in $\mathbb{S}$.

What if we want to control the error more generally? We know that the projection onto the column space of $V$ with minimal error is $VV^\top$. We can control the error of the oblique projection in terms of this minimal error.

**Lemma 1** (Oblique Projection Error)**.** *The oblique projection error is bounded as*

$$\|g - V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top g\|_2 \leq \|(\mathbb{S}^\top V)^{-1}\|_2 \|g - VV^\top g\|_2.$$

*Proof.* Let $\Pi = V(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top$. Then

$$\|(I - \Pi)g\|_2 = \|(I - \Pi)(g - VV^\top g)\|_2 \leq \|I - \Pi\|_2 \|g - VV^\top g\|_2.$$

Noting

$$\|I - \Pi\|_2 = \|\Pi\|_2 \leq \|(\mathbb{S}^\top V)^{-1}\mathbb{S}^\top\|_2 = \|(\mathbb{S}^\top V)^{-1}\|_2$$

finishes the proof. □

Note that for a more quantitative bound relating the oblique projection error to the optimal projection error, we will need to focus on bounding $\|(\mathbb{S}^\top V)^{-1}\|_2$. Note that this is equivalent to showing that the smallest singular value of $\mathbb{S}^\top V$ is not too small. To show this we will introduce the following notation.

Given an $m \times n$ matrix $A$ and index sets

$$I = \{i_1, \ldots, i_k\}, \quad 1 \leq i_1 < i_2 < \cdots < i_k \leq m,$$
$$J = \{j_1, \ldots, j_\ell\}, \quad 1 \leq j_1 < j_2 < \cdots < j_\ell \leq n,$$

we let

$$A(I,J) = \begin{bmatrix} a_{i_1,j_1} & \dots & a_{i_1,j_\ell} \\ \vdots & & \vdots \\ a_{i_k,j_1} & \dots & a_{i_k,j_\ell} \end{bmatrix} \in \mathbb{R}^{k\times \ell}.$$

The full index set is denoted by :, e.g. $A(I,:)$.

Then we propose the index selection algorithm given in Algorithm 1.

---

**Algorithm 1** Simplified form of Gaussian elimination with column pivoting

---

**Input:** $n \times r$ matrix $U$
**Output:** "Good" index set $I \subset \{1,\dots,n\}$, $\#I = r$
Set $I = \emptyset$
**for** k $= 1$ to $r$ **do**
  Choose $i^* = \arg\max_{i=1,\dots,n} |u_{ik}|$.
  Set $I = I \cup \{i^*\}$.
  $U \leftarrow U - \frac{1}{u_{i^*k}} U(:,k)U(i^*,:)$
**end for**

---

First to get an intuition, we will walk through an example of Gaussian elimination without column pivoting (Algorithm 2) in case this is way of viewing elimination is new.

---

**Algorithm 2** Gaussian elimination (no pivoting) applied to $U \in \mathbb{R}^{n\times r}$

---

**for** k $= 1$ to $r$ **do**
  $L(:,k) \leftarrow \frac{1}{U_{kk}} U(:,k)$
  $R(k,:) \leftarrow U(k,:)$
  $U \leftarrow U - L(:,k)R(k,:)$
**end for**

---

Let's see an example of Algorithm 2:

1. Before iteration 1,

$$L = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$

2. Before iteration 2

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -3 & -8 \\ 0 & -4 & -8 \end{bmatrix}$$

3. Before iteration 3

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 3 & 4/3 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -8 \\ 0 & 0 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 8/3 \end{bmatrix}$$

4. After iteration 3
$$L = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -8 \\ 0 & 0 & 8/3 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So if we swap the rows in our matrix so that the column pivoting Algorithm 1 chooses indices in order $1, 2, 3, \ldots$, then we see that Algorithm 1 will result in a factorization
$$PU = LR,$$
where

- $P$ is a permutation matrix,
- $L = \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix}$ with $L_{11} \in \mathbb{R}^{r\times r}$ unit lower triangular and $\max L_{ij} \leq 1$,
- $R \in \mathbb{R}^{r\times r}$ is upper triangular.

We will use this factorization and the following lemma to prove a bound on the smallest singular value of the matrix resulting from the row selection algorithm.

**Lemma 2** (Theorem 8.15 in Higham, 2002)**.** *Let $T \in \mathbb{R}^{n\times n}$ be an upper triangular matrix satisfying*
$$|t_{ii}| > |t_{ij}| \quad \textit{for } j > i.$$
*Then*
$$1 \leq \min_i |t_{ii}| \cdot \|T^{-1}\|_2 \leq \frac{1}{3}\sqrt{4^n + 6n - 1} \leq 2^{n-1}.$$

**Theorem 3** (Row Selection Bound on Smallest Singular Value)**.** *For the index set returned by the greedy algorithm applied to orthonormal $U \in \mathbb{R}^{n\times r}$, it holds that*
$$\|U(I,:)^{-1}\|_2 \leq \sqrt{nr}2^{r-1}.$$

*Proof.* Assuming without loss of generality that $I = \{1, 2, \ldots, r\}$, we start from
$$PU = LR. \tag{4}$$
Partitioning $L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ with $L_1 \in \mathbb{R}^{r\times r}$, the factorization (4) implies
$$U(I,:) = L_1 R.$$
Because $PU$ is orthonormal, (4) also implies $\|R_1^{-1}\|_2 = \|L\|_2$ so that
$$\|U(I,:)^{-1}\|_2 \leq \|L_1^{-1}\|_2 \|R^{-1}\|_2 = \|L_1^{-1}\|_2 \|L\|_2.$$
Because the magnitudes of the entries of $L$ are bounded by $1$, we have
$$\|L\|_2 \leq \|L\|_F \leq \sqrt{nr} \max |L_{ij}| \leq \sqrt{nr}.$$
Applying Lemma 2 to $L_1^\top$ in order to bound $\|L_1^{-1}\|_2$ completes the proof. □

Back to DEIM. We saw that we wanted to bound $\|(\mathbb{S}^\top V)^{-1}\|_2$. The theorem we just proved gives us a bound if we choose $\mathbb{S}$ so that $\mathbb{S}^\top V = V(I,:)$ with $I$ coming from Algorithm 1. Thus we will take $V$ to be some left singular vectors of snapshots of our *nonlinear* term then apply Algorithm 1 to this basis. We summarize DEIM below:

1. Form the snapshot matrix of our nonlinearity:
$$G = \begin{bmatrix} g(u_h^0(\mu_1)) & \dots & g(u_h^{N_t}(\mu_1)) & \dots & g(u_h^0(\mu_p)) & \dots & g(u_h^{N_t}(\mu_p)) \end{bmatrix}.$$

2. Compute SVD of $G = V\Sigma Q^\top$ and truncate to the $r$th principal subspace of $G$ with $r' \ll N_x$ (i.e., retain the first $r'$ columns $V_{r'}$ of $V$).

3. Apply Algorithm 1 to $V_{r'}$ to obtain an index set $I = \{i_1, i_2, \dots, i_{r'}\}$, and let
$$\mathbb{S} = \begin{bmatrix} e_{i_1} & \dots & e_{i_{r'}} \end{bmatrix}.$$

4. Approximate
$$g(u) \approx V_{r'}(\mathbb{S}^\top V_{r'})^{-1}\mathbb{S}^\top g(u).$$

Since $g$ acts componentwise, this approximation only needs to compute $r'$ components of $g$, which is where the speedup comes from. Because in our POD we actually compute $U_r^\top V_{r'}(\mathbb{S}^\top V_{r'})^{-1}\mathbb{S}^\top g(u)$, we can store the matrix $U_r^\top V_{r'}(\mathbb{S}^\top V_{r'})^{-1} \in \mathbb{R}^{r\times r'}$, which is small, then we multiply this by the $r'$ selected rows of $g$, making the cost of evaluating this approximation $\mathcal{O}(rr')$ (assuming the $g$ is approximated in $\mathcal{O}(1)$ time). Thus, our final reduced order model for the Allen-Cahn equation is

$$\frac{\widetilde{u}^{n+1} - \widetilde{u}^n}{\Delta t} - U_r^\top A U_r \widetilde{u}^{n+1/2} = U_r^\top V_{r'}(\mathbb{S}^\top V_{r'})^{-1}\mathbb{S}^\top (U_r\widetilde{u}^n - (U_r\widetilde{u}^{n+1})^3).$$