Math 673

Multigrid Methods: A Mostly Matrix-Based Approach

Chapter 01: Classical Iterative Methods

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu — swise1@.utk.edu
The University of Tennessee

Fall 2024

Chapter 01, Part 1 of 2
Classical Iterative Methods

The Basic Linear System

In this chapter, we will discuss classical linear iterative methods for solving the linear system of equations

$$A\boldsymbol{u} = \boldsymbol{f}. \tag{1}$$

We will usually be concerned with linear systems that arise from the approximation of elliptic PDE problems, such as the poison problem

$$-\Delta u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega,$$

via the finite difference method (FDM) or finite element method (FEM). However, in this first chapter, we will not specifically take such a narrow view. Much of our discussion will remain at a general level. We only make the links to PDEs and discretization methods as applications of the general framework later in the book.

We will always assume that $A \in \mathbb{R}^{n\times n}$ is invertible, at least. For most of the text, we will assume that it is symmetric positive definite (SPD).

Notation and Basic Facts

Vectors and Vector Spaces

Ours is a matrix-based approach to multigrid. So, it makes sense to get some pesky notational issues for matrices and vectors clear from the start. We will assume familiarity and maturity with normed vector spaces.

We use the symbol $\mathbb{R}^n$ to denote the set of all vectors (or, more generally, n-tuples) of length n. This, of course, becomes a vector space with the usual definitions of vector addition and scalar multiplication. We define the set

$$\mathbb{R}^n_\star := \{\boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{u} \neq \boldsymbol{0}\}.$$

When we write $\boldsymbol{u} = [u_i] \in \mathbb{R}^n$, we will always understand this object as a column vector, or, equivalently, a matrix having n rows and 1 column. To access the elements of a vector, we write

$$[\boldsymbol{u}]_j = u_j, \quad j \in \{1, \ldots, n\}.$$

To simplify notation, we will usually write $\boldsymbol{u}$ instead of $\boldsymbol{u} = [u_i]$, with the understanding that the elements of a vector are represented by the same letter, but unbolded.

Matrices and Multiplication

To access the elements of a generic matrix $\mathrm{A} = [a_{i,j}] \in \mathbb{R}^{m \times n}$, we write

$$[\mathrm{A}]_{i,j} = a_{i,j}, \quad i \in \{1, \ldots, m\}, \quad j \in \{1, \ldots, n\}.$$

As with vectors, we will usually drop the $[a_{i,j}]$ part. Matrix-vector multiplication is represented as usual: for $\mathrm{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{u} \in \mathbb{R}^n$, the product $\mathrm{A}\boldsymbol{u}$ is a vector in $\mathbb{R}^m$ with components

$$[\mathrm{A}\boldsymbol{u}]_i = \sum_{j=1}^{n} a_{i,j} u_j, \quad i \in \{1, \ldots, m\}.$$

Matrix-matrix multiplication is also represented as usual: for $\mathrm{A} = [a_{i,j}] \in \mathbb{R}^{m \times p}$ and $\mathrm{B} = [b_{i,j}] \in \mathbb{R}^{p \times n}$, the product AB is a matrix in $\mathbb{R}^{m \times n}$ whose components are

$$[\mathrm{AB}]_{i,j} = \sum_{k=1}^{p} a_{i,k} b_{k,j}, \quad i \in \{1, \ldots, m\}, \quad j \in \{1, \ldots, n\}.$$

The Transpose Operation and Multiplication

Suppose that $A = [a_{i,j}] \in \mathbb{R}^{m \times n}$. The transpose of A, written A^T, is a matrix in $\mathbb{R}^{n \times m}$ with components

$$\left[A^T\right]_{i,j} = a_{j,i}, \quad i \in \{1, \ldots, n\}, \quad j \in \{1, \ldots, m\}.$$

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, the object

$$\boldsymbol{v}^T \boldsymbol{u} = \sum_{i=1}^{n} v_i u_i$$

is a real number. This definition can be viewed in the context of matrix vector multiplication.

$$\left[\boldsymbol{v}\boldsymbol{u}^T\right]_{i,j} = v_i u_j, \quad i, j \in \{1, \ldots, n\}$$

is a matrix in $\mathbb{R}^{n \times n}$, which can be deduced from matrix-matrix multiplication.

Definition

Define, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$(\boldsymbol{u}, \boldsymbol{v}) := \boldsymbol{v}^T \boldsymbol{u}.$$

This object $(\,\cdot\,,\,\cdot\,) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called the **canonical inner product** or **Euclidean inner product**. We say that $\mathrm{A} \in \mathbb{R}^{n\times n}$ is **symmetric** iff, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$(\mathrm{A}\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{u}, \mathrm{A}\boldsymbol{v})\,.$$

We say that A is **symmetric positive definite** iff it is symmetric and, for all $\boldsymbol{u} \in \mathbb{R}^n_\star := \mathbb{R}^n \setminus \{\mathbf{0}\}$,

$$\boldsymbol{u}^T \mathrm{A}\boldsymbol{u} = (\mathrm{A}\boldsymbol{u}, \boldsymbol{u}) > 0.$$

Suppose that $\mathrm{A} \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. Define, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$(\boldsymbol{u}, \boldsymbol{v})_{\mathrm{A}} := (\mathrm{A}\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{v}^T \mathrm{A}\boldsymbol{u}.$$

The reader can easily show the following:

Proposition

A matrix $\mathrm{A} \in \mathbb{R}^{n \times n}$ *is symmetric iff* $\mathrm{A} = \mathrm{A}^T$.

Proof.

Exercise. □

Now, there are many possible inner products on $\mathbb{R}^n$. It may be that a matrix is not symmetric in the usual sense but is symmetric with respect to some other inner product.

Definition

Suppose that $(\cdot,\cdot)_\star : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is an inner product, not necessarily the Euclidean inner product. We say that $A \in \mathbb{R}^{n\times n}$ is **symmetric with respect** $(\cdot,\cdot)_\star$ iff, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$(A\boldsymbol{u}, \boldsymbol{v})_\star = (\boldsymbol{u}, A\boldsymbol{v})_\star .$$

We say that A is **symmetric positive definite with respect** $(\cdot,\cdot)_\star$ iff it is symmetric with respect $(\cdot,\cdot)_\star$ and, for all $\boldsymbol{u} \in \mathbb{R}^n_\star := \mathbb{R}^n \setminus \{\boldsymbol{0}\}$,

$$(A\boldsymbol{u}, \boldsymbol{u})_\star > 0.$$

We will take advantage of the following keystone result.

Theorem (Spectral Decomposition Theorem)

If $A \in \mathbb{R}^{n \times n}$ *is symmetric, then all its eigenvalues are real. Moreover, there is an orthonormal basis (orthonormal with respect to the Euclidean inner product) of* $\mathbb{R}^{n \times n}$ *consisting of eigenvectors of* A.

Proof.

See Chapter 01 of Salgado and Wise (2023). □

This can be generalized, as you might imagine.

Theorem (General Spectral Decomposition Theorem)

Suppose that $(\cdot,\cdot)_\star : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is an inner product. If* $\mathrm{A} \in \mathbb{R}^{n\times n}$ *is symmetric with respect* $(\cdot,\cdot)_\star$*, then all its eigenvalues are real. Moreover, there is an orthonormal basis (orthonormal with respect* $(\cdot,\cdot)_\star$*) of* $\mathbb{R}^n$ *consisting of eigenvectors of* A*.*

Proof.

See Chapter 07 of Salgado and Wise (2023). □

Proposition

A matrix $\mathrm{A} \in \mathbb{R}^{n \times n}$ *is SPD iff* $\mathrm{A} = \mathrm{A}^T$ *and all of its eigenvalues are positive.*

Proof.

Exercise. □

The last result generalizes as well.

Proposition

Suppose that $(\cdot, \cdot)_\star : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is an inner product. A matrix* $\mathrm{A} \in \mathbb{R}^{n \times n}$ *is SPD with respect* $(\cdot, \cdot)_\star$ *iff* A *is symmetric with respect* $(\cdot, \cdot)_\star$ *and all of its eigenvalues are positive.*

Proof.

Exercise. □

Proposition

Suppose that $A, B \in \mathbb{R}^{n \times n}$ *are symmetric positive definite matrices. Then,* $(\cdot,\cdot)_A$ *and* $(\cdot,\cdot)_{B^{-1}}$ *are inner products on* $\mathbb{R}^n$*. Moreover, the product matrix* BA *is symmetric and positive definite, with respect to these inner products. Consequently, all of the eigenvalues of* BA *are positive, and there is an orthonormal basis (orthonormal with respect to either of the new inner products) of eigenvectors.*

Proof.

We leave it as an exercise for the reader to show that $(\cdot,\cdot)_A$ and $(\cdot,\cdot)_{B^{-1}}$ are inner products. At this point, the reader should re-familiarize himself or herself with the definition of inner product.

Proof (Cont.)

Let us show that the product matrix BA is SPD with respect to $(\cdot,\cdot)_{B^{-1}}$. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ be arbitrary. Then,

$$\begin{aligned}
(BA\boldsymbol{u}, \boldsymbol{v})_{B^{-1}} &= \boldsymbol{v}^T B^{-1} BA\boldsymbol{u} \\
&= \boldsymbol{v}^T A\boldsymbol{u} \\
&= \boldsymbol{v}^T ABB^{-1}\boldsymbol{u} \\
&= \boldsymbol{v}^T A^T B^T B^{-1}\boldsymbol{u} \\
&= (BA\boldsymbol{v})^T B^{-1}\boldsymbol{u} \\
&= (\boldsymbol{u}, BA\boldsymbol{v})_{B^{-1}} .
\end{aligned}$$

This shows that BA is symmetric with respect to $(\cdot,\cdot)_{B^{-1}}$.

Proof (Cont.)

Now, suppose that $\boldsymbol{u} \in \mathbb{R}^n_\star$. Then

$$(\mathrm{BA}\boldsymbol{u}, \boldsymbol{u})_{\mathrm{B}^{-1}} = \boldsymbol{u}^T \mathrm{B}^{-1} \mathrm{BA}\boldsymbol{u} = \boldsymbol{u}^T \mathrm{A}\boldsymbol{u} > 0.$$

This shows that BA is positive definite with respect to $(\cdot, \cdot)_{\mathrm{B}^{-1}}$.

The other points follow from the General Spectral Decomposition Theorem and some simple calculations. See Chapter 07 of Salgado and Wise (2023). □

This final result of the section, together with the last, suggests that SPD matrices and inner products are essentially the same objects, that is, they can be identified.

Proposition

Suppose that $(\cdot,\cdot)_\star : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is an inner product. Then there exists a unique SPD matrix* $\mathrm{A} \in \mathbb{R}^{n\times n}$ *such that, for all* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$(\mathrm{A}\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{u}, \boldsymbol{v})_\star .$$

Proof.

Exercise. □

General Linear Iterative Schemes (GLIS)

Definition (General Linear Iterative Scheme (GLIS))

Suppose that $A \in \mathbb{R}^{n\times n}$ is invertible, $\boldsymbol{f} \in \mathbb{R}^n$ is given, and $\boldsymbol{u} = A^{-1}\boldsymbol{f}$. Assume that $B \in \mathbb{R}^n$ is invertible and $\boldsymbol{u}^0 \in \mathbb{R}^n$ is given. A **general linear iterative scheme (GLIS)** is an algorithm for determining the terms of the sequence $\{\boldsymbol{u}^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ via the recursive iteration

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k + B(\boldsymbol{f} - A\boldsymbol{u}^k). \tag{2}$$

The matrix B is called the **iterator** of the GLIS. The **error sequence** is defined via

$$\boldsymbol{e}^k := \boldsymbol{u} - \boldsymbol{u}^k.$$

The **residual sequence** is defined via

$$\boldsymbol{r}^k := \boldsymbol{f} - A\boldsymbol{u}^k = A\boldsymbol{e}^k.$$

We say that the GLIS (2) **converges unconditionally** iff $\lim_{k\to\infty} \boldsymbol{e}^k = \boldsymbol{0}$, for all initial vectors $\boldsymbol{u}^0 \in \mathbb{R}^n$.

Remark

The application of B *should be simple and computationally inexpensive. In some sense* B *should approximate* A^{-1}*. In the extreme case that* $\mathrm{B} = \mathrm{A}^{-1}$*, the GLIS converges after one iteration.*

Remark

In the GLIS (2)*, the matrix* B *is often called a* **preconditioner** *of* A*, especially, when* A *and* B *are both symmetric positive definite (SPD). Suppose* $\boldsymbol{u}^0 = \mathbf{0}$*. Then*

$$\boldsymbol{u}^1 = \mathrm{B}\boldsymbol{f}.$$

Thus the action of the preconditioner B *on a vector* $\boldsymbol{f}$ *is equivalent to 1 iteration of the GLIS with* $\boldsymbol{u}^0 = \mathbf{0} \in \mathbb{R}^n$*. A GLIS (where* B *is SPD) may be used as (or viewed as) a preconditioner for a Krylov method, like the Conjugate Gradient method. A "good" preconditioner has the property that*

$$\kappa(\mathrm{BA}) := \frac{\lambda_n(\mathrm{BA})}{\lambda_1(\mathrm{BA})} = \mathcal{O}(1).$$

where

$$0 < \lambda_1(\mathrm{BA}) \leq \lambda_2(\mathrm{BA}) \leq \cdots \leq \lambda_n(\mathrm{BA}),$$

are the ordered positive eigenvalues of the matrix BA*.*

Note that this matrix is SPD with respect the inner products $(\cdot,\cdot)_{\mathrm{A}}$ *and* $(\cdot,\cdot)_{\mathrm{B}^{-1}}$*.* BA **is not typically even symmetric with respect to the usual Euclidean inner product.** *See Chapter 07 of Salgado and Wise (2023) for more details.*

The Fixed Point and The Error Equation

It is easy to see that $\boldsymbol{u} = \mathrm{A}^{-1}\boldsymbol{f}$ is a fixed point of the GLIS:

$$\boldsymbol{u} = \boldsymbol{u} + \mathrm{B}(\boldsymbol{f} - \mathrm{A}\boldsymbol{u}).$$

From the last equation, let us subtract Equation (2) to obtain

$$\boldsymbol{e}^{k+1} = \boldsymbol{e}^{k} - \mathrm{BA}\boldsymbol{e}^{k} = (\mathrm{I} - \mathrm{BA})\boldsymbol{e}^{k}.$$

The matrix $\mathrm{I} - \mathrm{BA}$ has an important role in the analysis of the GLIS. We give it a name.

Definition (Error Transfer Matrix)

Suppose that $A \in \mathbb{R}^{n\times n}$ is invertible, $\boldsymbol{f} \in \mathbb{R}^n$ is given, $\boldsymbol{u} = A^{-1}\boldsymbol{f}$, and $\boldsymbol{u}^k$ is the GLIS iteration at the k^{th} step. The **GLIS error transfer matrix** is defined as

$$T := I - BA. \tag{3}$$

It satisfies

$$\boldsymbol{e}^k = T\boldsymbol{e}^{k-1},$$

for each $k \in \mathbb{N}$, where $\boldsymbol{e}^k = \boldsymbol{u} - \boldsymbol{u}^k$ is the error vector.

The Spectral Convergence Theory

Definition (Convergent Matrix)

Let $C \in \mathbb{C}^{n\times n}$. C is called **convergent** iff for every $\epsilon > 0$, there exists a $K \in \mathbb{N}$, such that, if $k \geq K$, then, for all $1 \leq i, j \leq n$,

$$\left| \left[C^k \right]_{i,j} \right| \leq \epsilon,$$

where, here,

$$C^k = \underbrace{C \cdots C}_{k \text{ times}}.$$

Theorem

The following statements are equivalent

1. $\mathsf{C} \in \mathbb{C}^{n\times n}$ *is convergent.*
2. *For some induced matrix norm* $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}$,
$$\lim_{k\to\infty} \left\|\mathsf{C}^k\right\| = 0.$$
3. *For all induced matrix norms,*
$$\lim_{k\to\infty} \left\|\mathsf{C}^k\right\| = 0.$$
4. $\rho(\mathsf{C}) < 1$, *where* $\rho(\mathsf{C})$ *is the spectral radius of* C,
$$\rho(\mathsf{C}) = \max_{1\le i\le n} \{|\lambda_i| \mid \lambda_i \in \sigma(\mathsf{C})\},$$
and $\sigma(\mathsf{C})$ *is the spectrum of* $\mathsf{C} \in \mathbb{C}^{n\times n}$.
5. *For every* $\boldsymbol{x} \in \mathbb{C}^n$,
$$\lim_{k\to\infty} \mathsf{C}^k \boldsymbol{x} = \mathbf{0}.$$

Theorem (Convergence of GLIS)

Suppose that $\mathsf{A}, \mathsf{B} \in \mathbb{R}^{n\times n}$ *are invertible,* $\boldsymbol{f} \in \mathbb{R}^n$ *is given, and* $\boldsymbol{u} = \mathsf{A}^{-1}\boldsymbol{f}$. *The GLIS* (2) *converges unconditionally iff*

$$\rho(\mathsf{T}) < 1,$$

where $\mathsf{T} = \mathsf{I} - \mathsf{BA}$ *is the error transfer matrix.*

Proof.

($\Rightarrow$): Suppose (2) converges, for any $\boldsymbol{u}^0 \in \mathbb{C}^n$. (It must converge to $\boldsymbol{u} \in \mathbb{R}^n$, incidentally.) Then

$$\lim_{k\to\infty} \boldsymbol{e}^k = \boldsymbol{0}.$$

Let $(\lambda, \boldsymbol{w}) \in \mathbb{C} \times \mathbb{C}^n$ be an eigen-pair of T, with $\|\boldsymbol{w}\|_\infty = 1$, and set $\boldsymbol{e}^0 = \boldsymbol{w}$. (Note that we must resort to the complex vector space, generically, since a matrix with all real entries need not have real eigenvalues.) Then

$$\boldsymbol{e}^k = \mathsf{T}^k\boldsymbol{e}^0 = \lambda^k\boldsymbol{e}^0.$$

Proof (Cont.)

Thus

$$\left\| \boldsymbol{e}^k \right\|_\infty = |\lambda|^k \to 0.$$

This implies that $|\lambda| < 1$. Since $\lambda \in \sigma(\mathrm{T})$ was arbitrary,

$$\rho(\mathrm{T}) < 1.$$

($\Leftarrow$): If $\rho(\mathrm{T}) < 1$, then by the last theorem,

$$\lim_{k \to \infty} \mathrm{T}^k \boldsymbol{x} = \boldsymbol{0},$$

for any $\boldsymbol{x} \in \mathbb{C}^n$. Hence

$$\lim_{k \to \infty} \boldsymbol{e}^k = \lim_{k \to \infty} \mathrm{T}^k \boldsymbol{e}^0 = \boldsymbol{0},$$

and the proof is complete. □

Theorem (Sufficient Condition for Convergence)

If $\|\mathrm{T}\| < 1$, *for any induced matrix norm* $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}$, *then the GLIS* (2) *converges for any starting point* $\boldsymbol{u}^0 \in \mathbb{R}^n$. *Furthermore, we have the error estimates*

$$\left\|\boldsymbol{u} - \boldsymbol{u}^k\right\| \leq \|\mathrm{T}\|^k \left\|\boldsymbol{u} - \boldsymbol{u}^0\right\|$$

and

$$\left\|\boldsymbol{u} - \boldsymbol{u}^k\right\| \leq \frac{\|\mathrm{T}\|^k}{1 - \|\mathrm{T}\|} \left\|\boldsymbol{u}^1 - \boldsymbol{u}^0\right\|.$$

Proof.

We use the well-known fact that, for any induced matrix norm $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}$,

$$\rho(\mathrm{T}) \leq \|\mathrm{T}\|.$$

Thus, if $\|\mathrm{T}\| < 1$, it follows that $\rho(\mathrm{T}) < 1$. Using the last theorem, we get the desired convergence result.

Proof (Cont.)

Next, let us establish the error estimates. Using the consistency and sub-multiplicativity of the induced norm,

$$\begin{aligned} \left\| \boldsymbol{e}^k \right\| &= \left\| T^k \boldsymbol{e}^0 \right\| \\ &\leq \left\| T^k \right\| \left\| \boldsymbol{e}^0 \right\| \\ &\leq \|\mathsf{T}\|^k \left\| \boldsymbol{e}^0 \right\|, \end{aligned}$$

that is,

$$\left\| \boldsymbol{u} - \boldsymbol{u}^k \right\| \leq \|\mathsf{T}\|^k \left\| \boldsymbol{u} - \boldsymbol{u}^0 \right\|,$$

which is the first error estimate.

Proof (Cont.)

Now, write

$$\begin{aligned} \left\| \boldsymbol{e}^0 \right\| &= \left\| \boldsymbol{u} - \boldsymbol{u}^0 \right\| \\ &= \left\| \boldsymbol{u} - \boldsymbol{u}^1 + \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\| \\ &\leq \left\| \boldsymbol{u} - \boldsymbol{u}^1 \right\| + \left\| \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\| \end{aligned}$$

Since,

$$\boldsymbol{u} - \boldsymbol{u}^1 = \boldsymbol{e}^1 = \mathsf{T}\boldsymbol{e}^0 = \mathsf{T}(\boldsymbol{u} - \boldsymbol{u}^0),$$

it follows that

$$\left\| \boldsymbol{u} - \boldsymbol{u}^1 \right\| \leq \|\mathsf{T}\| \left\| \boldsymbol{u} - \boldsymbol{u}^0 \right\| = \|\mathsf{T}\| \left\| \boldsymbol{e}^0 \right\|.$$

Proof (Cont.)

Therefore,

$$\begin{aligned} \left\| \boldsymbol{e}^0 \right\| &= \left\| \boldsymbol{u} - \boldsymbol{u}^1 + \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\| \\ &\leq \left\| \boldsymbol{u} - \boldsymbol{u}^1 \right\| + \left\| \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\| \\ &\leq \|\mathrm{T}\| \left\| \boldsymbol{e}^0 \right\| + \left\| \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\|. \end{aligned}$$

Hence

$$\left\| \boldsymbol{e}^0 \right\| \leq \frac{1}{1 - \|\mathrm{T}\|} \left\| \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\|.$$

Substituting into the first estimate yields

$$\left\| \boldsymbol{e}^k \right\| \leq \|\mathrm{T}\|^k \left\| \boldsymbol{u} - \boldsymbol{u}^0 \right\| = \|\mathrm{T}\|^k \left\| \boldsymbol{e}^0 \right\| \leq \frac{\|\mathrm{T}\|^k}{1 - \|\mathrm{T}\|} \left\| \boldsymbol{u}^1 - \boldsymbol{u}^0 \right\|,$$

which is the desired second error estimate. □