Cimmino's Algorithm

Kate Suraev (s224854029)

October 4, 2025

Contents

Abstract

Cimmino's method is a row-action iterative algorithm that is used to solve and approximate solutions to high-dimensional and sparse linear systems, both consistent and inconsistent, where direct methods such as Gaussian elimination are computationally expensive or infeasible. Direct methods are *slow* for large systems, and often have substantial and impractical memory requirements. Cimmino's method offers an efficient alternative, capable of finding sufficiently accurate solutions at a viable computational cost. Its inherent parallelisability and low memory requirements make it particularly suitable for large-scale problems, such as those encountered in Computed Tomography (CT) image reconstruction.

This report provides an in-depth exploration of the theoretical foundations of Cimmino's algorithm. We present a formal proof of convergence, demonstrating that the iterative sequence converges to a solution of the system if it is consistent, or to the weighted least squares solution if it is inconsistent. The analysis is grounded in the orthogonal decomposition of $\mathbb{R}^n$ into the null space of the coefficient matrix and the range of its adjoint, culminating in the characterisation of the limit of the generated sequence and a practical demonstration of Cimmino's algorithm in CT image reconstruction.

1 Introduction

Cimmino's method generates a sequence of approximations $\{x^k\}$ by simultaneously reflecting the current approximation x^k across the hyperplanes defined by the rows of the coefficient matrix A to obtain a new approximation x^{k+1} as the weighted average of these reflections [1] Under certain mild conditions (outlined below), this sequence converges, regardless of whether the system $Ax = b$ is consistent or inconsistent [2].

The generated sequence $\{x^k\}$ converges to the weighted least squares solution of the system $Ax = b$ if it is inconsistent, or to one solution of $Ax = b$ if it is consistent. Our mathematical focus will be on the convergence conditions and properties of Cimmino's algorithm and the characterisation of its limit. To this end, we define the linear system that we are considering and introduce the main theorem that we will prove in this report.

Consider the linear system $Ax = b$, where A is an $m \times n$ coefficient matrix, b is an m-dimensional vector of constants, and x is an n-dimensional vector of variables. The system may be either consistent or inconsistent; however, $\text{rank}(A)$ must be at least 2 and the rows and columns of A must be non-zero [2]:

$$\text{rank}(A) \geq 2 \tag{1a}$$

$$A_i \neq 0 \quad \text{for } i = 1, \ldots, m \tag{1b}$$

$$A^j \neq 0 \quad \text{for } j = 1, \ldots, n \tag{1c}$$

where A_i is the i-th row of A and A^j is the j-th column of A.

Theorem 1.1 Convergence of Cimmino's Algorithm

Provided the conditions in Eqn. (1) are satisfied, then for any initial approximation $x^0 \in \mathbb{R}^n$, Cimmino's algorithm generates a sequence of approximations $\{x^k\}$ that converges to [2]:

(i) a solution of the system $Ax = b$ if the system is consistent, or

(ii) the weighted least squares solution $\min_x \|D(Ax - b)\|$ if it is inconsistent.

The proof of Theorem 1.1 relies on the analysis of the iterative structure of Cimmino's algorithm, specifically its interaction with the orthogonal decomposition of $\mathbb{R}^n$ into the null space of A and the range of its adjoint A^T. We will demonstrate that these interactions ensure convergence to the solutions specified in the theorem.

We begin by giving an overview of Cimmino's algorithm and defining the key components involved in its iterative process.

2 Cimmino's Algorithm

The rows A_i of the coefficient matrix A and the corresponding components b_i of the vector b define m hyperplanes in $\mathbb{R}^n$. Each hyperplane H_i is a subspace of dimension $(n-1)$ [3, Ch. 10][4] defined as:

$$H_i = \{x \mid A_i^T x = b_i\} \quad \text{for } i = 1, \ldots, m \tag{2}$$

where A_i is the i-th row of A and b_i is the i-th component of b.

In a consistent system, the solution lies at the intersection of all hyperplanes H_i. In an inconsistent system, no solution lies on all hyperplanes simultaneously; instead, we seek the solution that minimises the sum of squared distances to each hyperplane, which corresponds to the weighted least squares solution [2].

Each iteration of Cimmino's algorithm updates the approximation by reflecting the current approximation x^k across each hyperplane H_i. The reflection of an arbitrary vector across a hyperplane is a geometric transformation that produces a mirror image of the vector with respect to the hyperplane [5].

The reflection $r^{(k,i)}$ of x^k across a hyperplane H_i is given by:

$$r^{(k,i)} = x^k + 2\frac{b_i - A_i^T x^k}{\|A_i\|^2} A_i \quad \text{for } i = 1, \ldots, m \tag{3}$$

where A_i is the normal vector to the hyperplane H_i [3, Ch. 4] and $\|A_i\|$ is its Euclidean norm [2].

The subsequent approximation x^{k+1} is formed as a convex combination of the reflections $r^{(k,i)}$ across all hyperplanes H_i scaled by weights ω_i [1].

The weights ω_i are positive real numbers that satisfy the conditions [2]:

$$\omega_i > 0 \quad \text{for} \quad \text{and} \quad \omega = \sum_{i=1}^{m} \omega_i$$

where ω is the sum of the weights.

The rows A_i of A are scaled by the weights ω_i to adjust the influence of each hyperplane in the computation of the next approximation, forming a diagonal matrix D with the scaled weights on its diagonal [2].

$$D = \begin{bmatrix} \frac{\sqrt{\omega_i}}{\|A_1\|} & 0 & \dots & 0 \\ 0 & \frac{\sqrt{\omega_2}}{\|A_2\|} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sqrt{\omega_m}}{\|A_m\|} \end{bmatrix} \tag{4}$$

As a result, the next approximation x^{k+1} is computed as [2]:

$$x^{k+1} = \sum_{i=1}^{m} \frac{\omega_i}{\omega} r^{(k,i)} = x^k + 2 \sum_{i=1}^{m} \frac{\omega_i}{\omega} \frac{b_i - A_i^T x^k}{\|A_i\|^2} A_i \quad \text{for } k = 0, 1, 2, \dots \tag{5}$$

where x^0 is an arbitrary initial approximation or guess. The expression in Eqn. (5) is elegantly rewritten in matrix form as [2]:

$$x^{k+1} = x^k + \frac{2}{\omega} A^T D^T D (b - Ax^k) \quad \text{for } k = 0, 1, 2, \dots \tag{6}$$

We note that for a particular choice of weights, specifically $\omega_i = \|A_i\|^2$ for all i [2],

$$\frac{\sqrt{\omega_i}}{\|A_i\|} = \frac{\sqrt{\|A_i\|^2}}{\|A_i\|} = 1 \quad \text{for } i = 1, \dots, m \tag{7}$$

leads to the result of $D = I$, the identity matrix, in Eqn. (4) and hence D plays no role in the computation of the next approximation x^{k+1} in Eqn. (6).

The behaviour we observe is that Cimmino's algorithm yields an approximation x^{k+1} that is strictly closer to the solution x^* (particular or weighted least squares) than the current approximation x^k [2]. Formally,

$$\|x^{k+1} - x^*\| < \|x^k - x^*\| \tag{8}$$

ensuring that each iteration of the algorithm refines the approximation towards a solution in Theorem 1.1.

At first glance, Cimmino's algorithm appears straightforward and applying it in practice is not particularly complex. However, the underlying mathematical principles that guarantee its convergence and effectiveness are intricate and require a deep understanding of linear algebra. The discussions throughout this report will explore the validity of Eqn. (8) and the conditions under which it holds. Before delving into the convergence analysis, we establish some foundational concepts, definitions and results in the realm of linear operators, orthogonal subspaces of $\mathbb{R}^n$, and vector decompositions. These concepts will be instrumental in understanding the behaviour of the algorithm as $k \to \infty$ and the nature of its limit.

3 Foundations for the Convergence of Cimmino's Method

To study the convergence properties of Cimmino's method, we rewrite the iterative formula in Eqn. (5) in terms of linear operators T and R as follows [2]:

$$x^{k+1} = Tx^k + Rb \quad \text{for } k = 0, 1, 2, \ldots \tag{9}$$

where

$$T := I - \frac{2}{\omega} A^T D^T D A \tag{10}$$

$$R := \frac{2}{\omega} A^T D^T D. \tag{11}$$

Here, T describes the transformation applied to the current approximation x^k, while R describes the contribution of the constant vector b to the next approximation.

The convergence of the generated sequence $\{x^k\}$ depends on the properties of the linear operators T and R.

To analyse the convergence of the sequence $\{x^k\}$, we examine the action of T and R on any vector in $\mathbb{R}^n$. Specifically, we will uniquely decompose $\mathbb{R}^n$ into the direct sum of the null space of A and the range of A^T:

$$\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A^T)$$

This decomposition allows us to analyse the behaviour of T and R on each subspace independently, providing insights into the convergence of the algorithm.

Before proceeding, we introduce key definitions and theorems that will be crucial in our analysis.

3.1 Linear Operators, Null Space, and Range

Definition 3.1 Linear operator

A linear operator $T : V \to W$ is a function mapping vectors from one vector space V to another vector space W while preserving vector addition and scalar multiplication [3, Ch. 7]. More formally, for all vectors $u, v \in V$ and scalar $k \in \mathbb{R}$, the following properties hold:

$$T(u + v) = T(u) + T(v)$$
$$T(ku) = kT(u)$$

Definition 3.2 Null space of a linear operator

The null space of a linear operator $T : V \to W$, denoted $\mathcal{N}(T)$, is the set of all vectors $v \in V$ for which $T(v) = 0$ [3, Ch. 7].

$$\mathcal{N}(T) = \{v \in V \mid T(v) = 0\}$$

Definition 3.3 Range of a linear operator

The range of a linear operator $T : V \to W$, denoted $\mathcal{R}(T)$, is the set of all vectors $w \in W$ that can be written as $w = T(v)$ for some $v \in V$ [3, Ch. 7].

$$\mathcal{R}(T) = \{T(v) \mid v \in V\}$$

We first verify that the null space and the range of a linear operator are subspaces of the respective vector spaces. Specifically, that they satisfy the properties of a subspace: contain the zero vector, are closed under addition, and are closed under scalar multiplication [3, Ch. 5].

Theorem 3.4 Null space and range are subspaces

Let $T : V \to W$ be a linear operator. Then, the null space $\mathcal{N}(T)$ and the range $\mathcal{R}(T)$ are subspaces of the vector spaces, V and W, respectively [3, Ch. 7].

To show that $\mathcal{N}(T)$ is a subspace of V, we need to verify the three properties of a subspace [3, Ch. 5].

1. Contains the zero vector: $0 \in V$ satisfies $T(0_V) = 0_W$, hence $0 \in \mathcal{N}(T)$.
2. Closed under addition: Let $u, v \in \mathcal{N}(T)$, then $T(u) = 0$ and $T(v) = 0$. Therefore,
$$T(u + v) = T(u) + T(v) = 0 + 0 = 0$$
So, $u + v \in \mathcal{N}(T)$.
3. Closed under scalar multiplication: Let $v \in \mathcal{N}(T)$ and k be a scalar. Then,
$$T(kv) = kT(v) = \langle k, 0\rangle = 0$$
So, $kv \in \mathcal{N}(T)$.

Similarly, we verify that $\mathcal{R}(T)$ is a subspace of W.

1. Contains the zero vector: $T(0_V) = 0_W$, hence $0_W \in \mathcal{R}(T)$.
2. Closed under addition: Let $w_1, w_2 \in \mathcal{R}(T)$, then there exist $v_1, v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Therefore,
$$T(v_1 + v_2) = T(v_1) + T(v_2) = w_1 + w_2$$
So, $w_1 + w_2 \in \mathcal{R}(T)$.
3. Closed under scalar multiplication: Let $w \in \mathcal{R}(T)$ and k be a scalar. We need to show that $kw \in \mathcal{R}(T)$. Since $w \in \mathcal{R}(T)$, there exists a vector $v \in V$ such that $T(v) = w$. Therefore,
$$T(kv) = kT(v) = kw$$
So, $kw \in \mathcal{R}(T)$.

We have shown that both $N(T)$ and $\mathcal{R}(T)$ satisfy the properties of a subspace, hence they are subspaces of V and W, respectively. Now we establish some important results about the null space and the range of a linear operator. Our goal is to show that the orthogonal complement of the null space of a linear operator is equal to the range of its adjoint operator, i.e., $\mathcal{N}(T)^{\perp} = \mathcal{R}(T^*)$.

3.2 Orthogonal Complements

Definition 3.5 Orthogonal complement

Given a finite dimensional inner product space V, let S be a subspace of V. The orthogonal complement of S, denoted by $S^{\perp}$, consists of all vectors in V that are orthogonal to every vector in S [3, Ch. 8]. More formally,

$$S^{\perp} = \{x \in V : \langle x, s \rangle = 0 \text{ for all } s \in S\}$$

Theorem 3.6 Orthogonal complement of a subspace is a subspace

If S is a subspace of a finite dimensional inner product space V, then the orthogonal complement $S^{\perp}$ is also a subspace of V [3, Ch. 6]

As before, we show that $S^{\perp}$ satisfies the three properties of a subspace.

1. Contains the zero vector: $0 \in V$ is orthogonal to every vector in S since, for any $s \in S$, $\langle 0, s \rangle = 0$. Therefore, $0 \in S^{\perp}$.
2. Closed under addition: Let $x, y \in S^{\perp}$. Then, for any $s \in S$, we have $\langle x, s \rangle = 0$ and $\langle y, s \rangle = 0$. Therefore,

$$\langle x + y, s \rangle = \langle x, s \rangle + \langle y, s \rangle = 0 + 0 = 0$$

   So, $x + y \in S^{\perp}$.
3. Closed under scalar multiplication: Let $x \in S^{\perp}$ and k be a scalar. We want to show that $kx \in S^{\perp}$. For any $s \in S$, $\langle x, s \rangle = 0$. Therefore,

$$\langle kx, s \rangle = k\langle x, s \rangle = \langle k, 0 \rangle = 0$$

   So, $kx \in S^{\perp}$.

Since $S^{\perp}$ satisfies all three properties, it is a subspace of V, and Theorem 3.6 is proven.

Theorem 3.7 Double orthogonal complement of a subspace is the subspace itself

Let S be a subspace of a finite dimensional inner product space V. Then, the double orthogonal complement of S, denoted $S^{\perp\perp}$, is equal to S itself. That is, $S^{\perp\perp} = S$ [3, Ch. 8].

To prove that $S^{\perp\perp} = S$, we need to show that (i) $S^{\perp\perp} \subseteq S$ and (ii) $S \subseteq S^{\perp\perp}$.

(i) $S^{\perp\perp} \subseteq S$

Let $\{b_1, b_2, \ldots, b_k\}$ be an orthonormal basis for S, and let w be any vector in V.

Define the following:

$$u_i = \langle w, b_i \rangle b_i \quad \text{for } i = 1, 2, \ldots, k \tag{12}$$

where u_i is the projection of w onto the basis vector b_i and $u = u_1 + u_2 + \ldots + u_k = \sum_{i=1}^{k} \langle w, b_i \rangle b_i$ is the linear combination of the basis vectors b_i with coefficients $\langle w, b_i \rangle$, thus, $u \in S$, and

$$v = w - u \tag{13}$$

where v is the vector w minus its orthogonal projection u onto S.

We want to show that v is orthogonal to every basis vector b_i of S for all $i = 1, \ldots, k$:

For every basis vector b_i of S,

$$\langle v, b_i \rangle = \langle w - u, b_i \rangle = \langle w, b_i \rangle - \langle u, b_i \rangle = \langle w, b_i \rangle - \sum_{j=1}^{k} \langle w, b_j \rangle \langle b_j, b_i \rangle$$

Since the basis vectors b_i are orthonormal, $\langle b_j, b_i \rangle = 0$ for $j \neq i$ and $\langle b_i, b_i \rangle = 1$ [3, Ch. 10]. Therefore, the only term that does not vanish in the sum is when $j = i$, which gives us:

$$\langle v, b_i \rangle = \langle w, b_i \rangle - \langle w, b_i \rangle = 0$$

Therefore, v is orthogonal to every b_i in S because, for any $x \in S$, we have:

$$\langle v, x \rangle = \langle v, \sum_{i=1}^{k} x_i b_i \rangle = \sum_{i=1}^{k} x_i \langle v, b_i \rangle = \sum_{i=1}^{k} x_i (0) = 0$$

Thus, $v \in S^{\perp}$.

If we rearrange Eqn. (13), we get $w = u + v$ with $u \in S$ and $v \in S^{\perp}$.

We derive an intermediate result: any vector $w \in V$ can be written as the sum of two vectors, $u \in S$ and $v \in S^{\perp}$. Here, u represents the orthogonal projection of w onto S, while v is the orthogonal projection of w onto $S^{\perp}$.

$$w = u + v = P_S w + P_{S^{\perp}} w \tag{14}$$

where P is the projection operator such that $P_S w \in S$ and $P_{S^{\perp}} w \in S^{\perp}$.

Now returning to our main proof of Theorem 3.7.

Suppose $w \in S^{\perp\perp}$. By definition, w is orthogonal to every vector v in $S^{\perp}$.

$$\langle w, v \rangle = 0 \quad \text{for all } v \in S^{\perp} \tag{15}$$

Also, since $u \in S$ and $v \in S^{\perp}$,

$$\langle u, v \rangle = 0 \quad \text{for all } v \in S^{\perp} \tag{16}$$

We want to show that $w \in S^{\perp\perp}$ also implies $w \in S$. Using Eqn. 15 and Eqn. 16 and the fact that $v = w - u$,

$$\langle w, v \rangle - \langle u, v \rangle = \langle w - u, v \rangle = \langle v, v \rangle = \|v\|^2 = 0$$

Since $\|v\|^2 = 0$, this implies $v = 0$ [3, Ch. 10]. Therefore, $w = u + v = u + 0 = u \in S$ and therefore, $S^{\perp\perp} \subseteq S$.

(ii) $S \subseteq S^{\perp\perp}$

Let $x \in S$. By definition, x is orthogonal to every vector in $S^{\perp}$. However, x is in the orthogonal complement of $S^{\perp}$, i.e., $x \in S^{\perp\perp}$. Therefore, $S \subseteq S^{\perp\perp}$.

Having shown both inclusions, we conclude that $S^{\perp\perp} = S$.

From this proof of Theorem 3.7, we derived an important intermediate result in Eqn. (14), that any vector w in V can be expressed as the sum of a vector in S and a vector in $S^{\perp}$. We now extend this result to show that V can be uniquely decomposed into the direct sum of S and $S^{\perp}$.

3.3 Direct Sum Decomposition

Definition 3.8 Direct Sum Decomposition

Let U and W be subspaces of a vector space V. Then, V is a direct sum of U and W if

$$V = U + W \quad \text{and} \quad U \cap W = \{0\}$$

where $U + W = \{u + w : u \in U, w \in W\}$. Then, we write $V = U \oplus W$, where W is called a complement of U in V [3, Ch. 9].

Theorem 3.9 Unique Decomposition into Direct Sum

Let S be a subspace of a finite dimensional inner product space V. Then, V has a unique decomposition into the direct sum of S and its orthogonal complement $S^{\perp}$. That is,

$$V = S \oplus S^{\perp}$$

To show that $V = S \oplus S^{\perp}$, we verify that it satisfies the two conditions in Definition 3.8.

$(V = S + S^{\perp})$ This follows directly from Eqn. (14): $w = u + v$ where $u \in S$ and $v \in S^{\perp}$.

$(S \cap S^{\perp} = \{0\})$ Let $x \in S \cap S^{\perp}$. Then, $x \in S$ and $x \in S^{\perp}$. Since $x \in S^{\perp}$, x must be orthogonal to every vector in S, including itself. Therefore, $\langle x, x\rangle = \|x\|^2 = 0$, which implies that $x = 0$ [3, Ch. 10] and hence, $S \cap S^{\perp} = \{0\}$.

To show that this decomposition is unique, suppose a vector $w \in V$ can be written in two distinct ways [6]:

$$w = s_1 + s_1^{\perp} = s_2 + s_2^{\perp} \text{ where } s_1, s_2 \in S \text{ and } s_1^{\perp}, s_2^{\perp} \in S^{\perp}$$

Subtracting the two expressions for w, we get:

$$(s_1 - s_2) + (s_1^{\perp} - s_2^{\perp}) = 0$$

Since $S \cap S^{\perp} = \{0\}$, it follows that $s_1 - s_2 = 0$ and $s_1^{\perp} - s_2^{\perp} = 0$ [6]. Therefore, $s_1 = s_2$ and $s_1^{\perp} = s_2^{\perp}$, confirming the uniqueness of the direct sum decomposition, $V = S \oplus S^{\perp}$.

3.4 Null Space and Range of the Adjoint Operator

Lemma 1. *Let A and B be two subspaces of a finite dimensional inner product space V with $A \subseteq B$, then the orthogonal complement of B is a subset of the orthogonal complement of A, that is, $B^\perp \subseteq A^\perp$.*

Let $x \in B^\perp$. By definition of $B^\perp$, we have $\langle x, b\rangle = 0$ for all $b \in B$. Since $A \subseteq B$, it follows that $\langle x, a\rangle = 0$ for all $a \in A$. Therefore, $x \in A^\perp$, which implies that $B^\perp \subseteq A^\perp$, completing the proof of Lemma 1.

We now introduce the final theorem in this section that relates the null space of a linear operator to the range of its adjoint.

Definition 3.10 Adjoint Operator

The adjoint of a linear operator $T : V \to W$, denoted $T^* : W \to V$, is defined such that for all vectors $v \in V$ and $w \in W$, the following holds:

$$\langle T(v), w\rangle_W = \langle v, T^*(w)\rangle_V$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product defined in the spaces V and W [7].

Theorem 3.11 Orthogonal complement of the null space

For any linear operator T between two finite dimensional inner product spaces, the orthogonal complement of the null space of T is equal to the range of the adjoint operator T^*. That is,

$$\mathcal{N}(T)^\perp = \mathcal{R}(T^*)$$

We approach the proof of Theorem 3.11 by demonstrating the two inclusions: (i) $\mathcal{R}(T^*) \subseteq \mathcal{N}(T)^\perp$ and (ii) $\mathcal{N}(T)^\perp \subseteq \mathcal{R}(T^*)$.

(i) $\mathcal{R}(T^*) \subseteq \mathcal{N}(T)^\perp$

Let $x \in \mathcal{R}(T^*)$. By Definition 3.3, there exists a vector $w \in W$ such that $x = T^*(w)$.

Let $v \in \mathcal{N}(T)$. By Definition 3.2, $T(v) = 0$.

Using Definition 3.10 of the adjoint operator,

$$\langle x, v\rangle = \langle T^*(w), v\rangle = \langle w, T(v)\rangle = \langle w, 0\rangle = 0$$

Since $\langle x, v\rangle = 0$ for all $v \in \mathcal{N}(T)$, this implies that $x \in \mathcal{N}(T)^\perp$. Therefore, $\mathcal{R}(T^*) \subseteq \mathcal{N}(T)^\perp$.

(ii) $\mathcal{N}(T)^\perp \subseteq \mathcal{R}(T^*)$

In order to apply Lemma 1, we show $\mathcal{R}(T^*)^\perp \subset \mathcal{N}(T)$.

Let $x \in \mathcal{R}(T^*)^\perp$. By definition, x is orthogonal to every vector in $\mathcal{R}(T^*)$, in other words, $\langle x, y\rangle = 0$ for all $y \in \mathcal{R}(T^*)$. Since $y \in \mathcal{R}(T^*)$, by Definition 3.3, $y = T^*(w)$.

Again, using Definition 3.10 of the adjoint operator, we have:

$$\langle x, y \rangle = \langle x, T^*(w) \rangle = \langle T(x), w \rangle = 0$$

must hold for all $w \in W$.

The statement $\langle T(x), w \rangle = 0$ for all $w \in W$ means that $T(x)$ is orthogonal to every vector in W. The only vector with this property is the zero vector, so $T(x) = 0$.

$T(x) = 0$ implies that $x \in \mathcal{N}(T)$. Therefore, we have shown that $\mathcal{R}(T^*)^\perp \subseteq \mathcal{N}(T)$.

Finally, let $A = \mathcal{R}(T^*)^\perp$ and $B = \mathcal{N}(T)$. By Theorems 3.4 and 3.6 $\mathcal{R}(T^*)^\perp$ and $\mathcal{N}(T)$ are subspaces of V, From our proof that $\mathcal{R}(T^*)^\perp \subseteq \mathcal{N}(T)$, we can conclude that $A \subseteq B$. Using Lemma 1, since $A \subseteq B$, we have $B^\perp \subseteq A^\perp$, i.e., $\mathcal{N}(T)^\perp \subseteq \mathcal{R}(T^*)^{\perp\perp}$.

By Theorem 3.7, $\mathcal{R}(T^*)^{\perp\perp} = \mathcal{R}(T^*)$. Therefore, $\mathcal{N}(T)^\perp \subseteq \mathcal{R}(T^*)$.

Since we have shown both $\mathcal{R}(T^*) \subseteq \mathcal{N}(T)^\perp$ and $\mathcal{N}(T)^\perp \subseteq \mathcal{R}(T^*)$, we can conclude that $\mathcal{N}(T)^\perp = \mathcal{R}(T^*)$, completing the proof of Theorem 3.11. Therefore, by Theorem 3.9, we have a unique decomposition of V into the direct sum of $\mathcal{N}(T)$ and $\mathcal{R}(T^*)$:

$$V = \mathcal{N}(T) \oplus \mathcal{R}(T^*) \tag{17}$$

With the unique decomposition $V = \mathcal{N}(T) \oplus \mathcal{R}(T^*)$ established, we can now analyse the behaviour of Cimmino's algorithm on each component of the decomposition separately. By rewriting Cimmino's algorithm in terms of the linear operators T and R, we can study how the algorithm iteratively refines the approximation x^k towards a solution and its behaviour on the orthogonal components of the vector space $\mathbb{R}^n$.

4 Geometric Interpretation of Cimmino's Algorithm

Recall from Eqn. (9) that we have rewritten Cimmino's algorithm (Eqn. 6) as follows:

$$x^{k+1} = Tx^k + Rb$$

where

$$T := I - \frac{2}{\omega} A^T D^T D A \quad \text{and} \quad R := \frac{2}{\omega} A^T D^T D$$

We note that T can be expressed as the sum of m orthogonal reflectors S_i, also known as Householder transformations [2], [8]:

$$T = \sum_{i=1}^{m} S_i \tag{18}$$

where

$$S_i = I - \frac{2}{\omega} \frac{A_i A_i^T}{\|A_i\|^2} \tag{19}$$

We first want to verify that S_i indeed reflects a vector x across the hyperplane orthogonal to A_i.

4.1 Orthogonal Reflectors

Lemma 2. *Given a vector $x \in \mathbb{R}^n$, the transformation S_i, defined as*

$$S_i = I - \frac{2}{\omega} \frac{A_i A_i^T}{\|A_i\|^2}$$

reflects x across the hyperplane orthogonal to a non-zero vector $A_i \in \mathbb{R}^n$. Furthermore, S_i is an orthogonal matrix.

Let $\omega = 1$. Then, S_i simplifies to:

$$S_i = I - 2\frac{A_i A_i^T}{\|A_i\|^2}$$

Let H be the linear hyperplane orthogonal to A_i defined as [3, Ch. 8]:

$$H = \{x : A_i^T x = 0\}$$

The orthogonal complement of H, denoted by $H^\perp$, is a subspace spanned by the normal vector A_i [3, Ch. 4]. The projection of x onto $H^\perp$ is given by:

$$P_{H^\perp} x = \frac{A_i^T x}{\|A_i\|^2} A_i$$

This projection gives us the component of x that lies in the direction of A_i.

From our proof of Theorem 3.11, we know that any vector x in $\mathbb{R}^n$ can be uniquely decomposed into two components: one in H and one in $H^\perp$. Specifically, we can write:

$$x = P_H x + P_{H^\perp} x$$

By rearranging the equation, we can express the projection of x onto H as:

$$P_H x = x - P_{H^\perp} x = x - \frac{A_i^T x}{\|A_i\|^2} A_i$$

Then, the reflected point r is found by starting at the projection $P_H x$ and adding the same displacement vector that leads from the original point x to its projection [5]:

$$r = P_H x + (P_H x - x) = 2P_H x - x$$

Substituting $P_H x$ into the equation for r, we get:

$$r = 2\left(x - \frac{A_i^T x}{\|A_i\|_2^2} A_i\right) - x = x - 2\frac{A_i^T x}{\|A_i\|_2^2} A_i$$

Applying S_i to a vector x gives us:

$$S_i x = \left(I - 2\frac{A_i A_i^T}{\|A_i\|_2^2}\right) x = Ix - 2\frac{(A_i A_i^T) x}{\|A_i\|_2^2} = x - 2\frac{A_i (A_i^T x)}{\|A_i\|_2^2} = x - 2\frac{A_i^T x}{\|A_i\|_2^2} A_i \quad (20)$$

The result of applying x to S_i is the same as the formula for the reflected point r. Therefore, the linear transformation S_i reflects a vector $x \in \mathbb{R}^n$ across the hyperplane orthogonal to A_i.

To show that S_i is orthogonal, we need to verify that $S_i^T S_i = I$ [9]. First, we compute the transpose of S_i:

$$S_i^T = \left(I - 2\frac{A_i A_i^T}{\|A_i\|_2^2}\right)^T = I^T - 2\left(\frac{A_i A_i^T}{\|A_i\|_2^2}\right)^T = I - 2\frac{(A_i A_i^T)^T}{\|A_i\|_2^2} = I - 2\frac{A_i A_i^T}{\|A_i\|_2^2} = S_i$$

Given that $S_i^T = S_i$, we see that S_i is symmetric and compute the product $S_i^T S_i$ as follows:

$$\begin{aligned} S_i^T S_i &= \left(I - 2\frac{A_i A_i^T}{\|A_i\|_2^2}\right)^2 \\ &= I^2 - 2I\frac{A_i A_i^T}{\|A_i\|_2^2} - 2\frac{A_i A_i^T}{\|A_i\|_2^2}I + 4\left(\frac{A_i A_i^T}{\|A_i\|_2^2}\right)^2 \\ &= I - 4\frac{A_i A_i^T}{\|A_i\|_2^2} + 4\frac{A_i(\|A_i\|_2^2)A_i^T}{\|A_i\|_2^4} \\ &= I - 4\frac{A_i A_i^T}{\|A_i\|_2^2} + 4\frac{A_i A_i^T}{\|A_i\|_2^2} \\ &= I \end{aligned}$$

Since $S_i^T S_i = I$, we conclude that S_i is an orthogonal matrix and, as a result, T is the sum of m orthogonal matrices.

4.2 Invariant Subspaces

In this section, we analyse the behaviour of the linear operator T on the null space of A, $\mathcal{N}(A)$, and the range of A^T, $\mathcal{R}(A^T)$. We first establish that $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$ are invariant under T, meaning that applying T to any vector in these subspaces results in a vector in the same subspace. Similarly, we show that the operator R maps any vector in $\mathbb{R}^m$ into $\mathcal{R}(A^T)$.

Furthermore, we study how T and R act on these subspaces. Specifically, we demonstrate that T preserves the norm of vectors in $\mathcal{N}(A)$ and acts as a contraction on $\mathcal{R}(A^T)$, while R maps vectors into $\mathcal{R}(A^T)$ in a way that complements the action of T. These properties allow us to analyse T and R separately on $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$, simplifying the study of their overall behaviour on $\mathbb{R}^n$.

Definition 4.1 Invariant Subspace

Let $T : V \to W$ be a linear operator. A subspace U of V such that $U \subseteq V$ is called an invariant subspace under T, or T-invariant, if $T(U) \subseteq U$. In other words, for every vector $u \in U$, $T(u) \in U$ [3, Ch. 9].

Lemma 3. *Given the linear operators T and R defined as*

$$T := I - \frac{2}{\omega}A^T D^T DA \quad \text{and} \quad R := \frac{2}{\omega}A^T D^T D$$

the following properties hold [2]:

(i) $\mathcal{N}(A)$ *is an invariant subspace of T: if* $x \in \mathcal{N}(A)$*, then* $T(x) = x \in \mathcal{N}(A)$*.*

(ii) $\mathcal{R}(A^T)$ *is an invariant subspace of T: if* $x \in \mathcal{R}(A^T)$*, then* $T(x) \in \mathcal{R}(A^T)$*.*

(iii) $\mathcal{R}(A^T)$ *is an invariant subspace of R: For any* $z \in \mathbb{R}^m$*,* $Rz \in \mathcal{R}(A^T)$*.*

(i) By Definition 3.2 of the null space, for any $x \in \mathcal{N}(A)$, we have $Ax = 0$. We want to show that $T(x) \in \mathcal{N}(A)$.

$$T(x) = \left(I - \frac{2}{w}A^T D^T DA\right)x = x - \frac{2}{w}A^T D^T DAx = x - \frac{2}{w}A^T D^T D(0) = x$$

Since $x \in \mathcal{N}(A)$, this implies that $T(x) = x \in \mathcal{N}(A)$ and hence, $\mathcal{N}(A)$ is an invariant subspace of T.

(ii) Let $x \in \mathcal{R}(A^T)$, then, by Definition 3.3, there exists a vector $z \in \mathbb{R}^m$ such that $x = A^T z$. We want to show that $T(x) \in \mathcal{R}(A^T)$.

$$T(x) = \left(I - \frac{2}{w}A^T D^T DA\right)x = x - \frac{2}{w}A^T D^T DAx = A^T z - \frac{2}{w}A^T D^T DAA^T z$$

Factoring out the A^T term:

$$T(x) = A^T\left(z - \frac{2}{w}D^T DAA^T z\right)$$

Let $v = z - \frac{2}{w}D^T DAA^T z$. Since $z \in \mathbb{R}^m$, and $D^T DAA^T z$ is a linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^m$, it follows that $v \in \mathbb{R}^m$. Then, $T(x) = A^T v$, which by Definition 3.3, implies that $T(x) \in \mathcal{R}(A^T)$. Therefore, $\mathcal{R}(A^T)$ is an invariant subspace of T.

(iii) Let $z \in \mathbb{R}^m$. We want to show that $Rz \in \mathcal{R}(A^T)$. First applying the linear operator R to z:

$$Rz = \frac{2}{w}A^T D^T Dz$$

We can again factor out the A^T term:

$$Rz = A^T\left(\frac{2}{w}D^T Dz\right)$$

Let $u = \frac{2}{w}D^T Dz$. Since $z \in \mathbb{R}^m$, and $D^T D$ is a linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^m$, it follows that $u \in \mathbb{R}^m$. Then, $Rz = A^T u$, which by Definition 3.3, implies that $Rz \in \mathcal{R}(A^T)$. Therefore, $\mathcal{R}(A^T)$ is an invariant subspace of R, ensuring that applying R to any vector in $\mathbb{R}^m$ results in a vector that lies within $\mathcal{R}(A^T)$.

We have established that $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$ are invariant subspaces of the linear operator T and that R maps any vector in $\mathbb{R}^m$ into $\mathcal{R}(A^T)$. This brings us to the next set of properties we aim to establish about the operator T. In particular, we will show that T is non-expansive in $\mathbb{R}^n$, meaning it does not increase the norm of any vector it is applied to. Additionally, we want to extend this property to show that T acts as a contraction on $\mathcal{R}(A^T)$, strictly decreasing the norm of any non-zero vector in this subspace.

4.3 Non-expansiveness and Contraction Properties

Lemma 4. *Given the linear operator T defined as*

$$T := I - \frac{2}{\omega} A^T D^T D A$$

and the orthogonal transformations S_i defined as

$$S_i := I - \frac{2}{\omega} \frac{A_i A_i^T}{\|A_i\|^2}$$

then, the following properties hold [2]:

(i) For any $x \in \mathbb{R}^n$, the transformations S_i preserve the norm of x: $\|S_i x\| = \|x\|$. Consequently, $\|S_i\| = 1$ for all $i = 1, \ldots, m$.

(ii) T has an operator norm of 1, $\|T\| = 1$.

(iii) If $\text{rank}(A) \geq 2$, then the restriction of T to the subspace $\mathcal{R}(A^T)$, denoted $T\mid_{\mathcal{R}(A^T)}$, is a contraction, satisfying $\|T\mid_{\mathcal{R}(A^T)}\| < 1$.

(iv) $\|T(x)\| = \|x\| \iff x \in \mathcal{N}(A)$.

(i) From Lemma 2, we know that S_i is an orthogonal matrix, specifically $S_i^T S_i = I$. We now want to show that S_i preserves the norm of any vector $x \in \mathbb{R}^n$ and, as a result, the norm of S_i is 1.

Using the fact that the squared norm of a vector v can be expressed as $\|v\|^2 = v^T v$ [3, Ch. 10] and $S_i^T S_i = I$, we compute the squared norm of $S_i x$ as follows:

$$\|S_i x\|^2 = (S_i x)^T S_i x = x^T S_i^T S_i x = x^T I x = x^T x = \|x\|^2$$

Since $\|S_i x\|^2 = \|x\|^2$, taking the square root of both sides gives $\|S_i x\| = \|x\|$ for any $x \in \mathbb{R}^n$. From this, we establish that S_i preserves the norm of any vector x it is applied to. Therefore, the operator norm of S_i for all $i = 1, \ldots, m$ is given by [10]:

$$\|S_i\| = \sup_{x \neq 0} \frac{\|S_i x\|}{\|x\|} = \sup_{x \neq 0} \frac{\|x\|}{\|x\|} = 1$$

Thus, $\|S_i\| = 1$ for all $i = 1, \ldots, m$, confirming that S_i is a norm-preserving transformation.

(ii) From Eqn. (18), we know that T can be expressed as the sum of m orthogonal matrices S_i. We will use this fact to show that the operator norm of T is 1.

For an arbitrary $x \in \mathbb{R}^n$, the action of T on x is given by:

$$Tx = \sum_{i=1}^{m} \frac{\omega_i}{\omega} S_i x$$

where $\frac{\omega_i}{\omega}$ are the non-negative weights that sum to 1. In other words, Tx is a convex combination of the vectors $S_i x$ for $i = 1, \ldots, m$ [11].

Geometrically, the vectors $S_i x$ lie on the same hypersphere of radius $\|x\|$ since $\|S_i x\| = \|x\|$ (Lemma 2 (i)) [2]. Then, $\|Tx\|$ will lie strictly inside the hypersphere unless all the vectors $S_i x$ are equal, in which case $\|Tx\| = \|x\|$, implying that $x \in \mathcal{N}(A)$.

Consider the following example, which illustrates this concept in $\mathbb{R}^2$. We define a 3×2 coefficient matrix A as follows:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 3 \end{bmatrix} \tag{21}$$

Let $x = (1, -3)$. Using Eqn. (20), we compute the reflected vectors $S_i x$ for $i = 1, 2, 3$. These reflections are plotted in Fig. 1 to illustrate the geometric behaviour of the transformations along with the convex combination Tx.

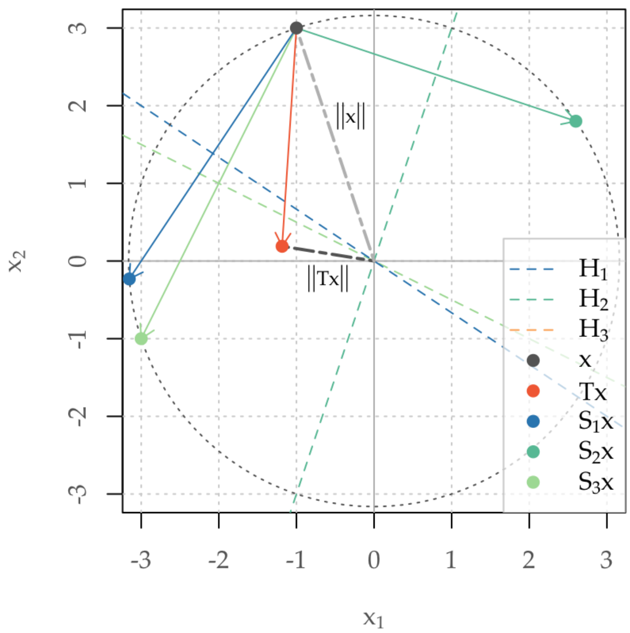


Figure 1: Eqn. (21) defines the linear hyperplanes H_i which x is reflected across. The vectors $S_i x$ lie on the hypersphere of radius $\|x\|$, centred at the origin. The vector Tx is the convex combination of the $S_i x$ and lies strictly inside the hypersphere, satisfying $\|Tx\| < \|x\|$. Drawn in R.

Our first goal in this proof is to show that $\|Tx\| \leq \|x\|$ holds with equality if and only if the vectors $S_i x$ are all equal. We will use the Cauchy-Schwarz inequality to establish this.

Definition 4.2 Cauchy-Schwarz Inequality

Let v and w be vectors in an inner product space. The Cauchy-Schwarz inequality states that:

$$|\langle v, w \rangle| \leq \|v\| \|w\| \tag{22}$$

with equality if and only if one of v or w is a scalar multiple of the other [3, Ch. 10].

Lemma 5. *Let x_1 and x_2 be two vectors in an inner product space, such that $\|x_1\| \leq \|x_2\|$, and let ω_1 and ω_2 be non-negative weights such that $\omega_1 + \omega_2 = 1$. Then*

$$\|\omega_1 x_1 + \omega_2 x_2\|^2 \leq \|x_2\|^2 \tag{23}$$

with equality if and only if $x_1 = x_2$.

We start by expanding the left-hand side of Eqn. (23) using the properties of inner products:

$$\begin{aligned}\|\omega_1 x_1 + w_2 x_2\|^2 &= \langle \omega_1 x_1 + w_2 x_2, \omega_1 x_1 + w_2 x_2 \rangle \\ &= \omega_1^2 \langle x_1, x_1 \rangle + w_2^2 \langle x_2, x_2 \rangle + 2\omega_1 w_2 \langle x_1, x_2 \rangle \\ &= \omega_1^2 \|x_1\|^2 + w_2^2 \|x_2\|^2 + 2\omega_1 w_2 \langle x_1, x_2 \rangle\end{aligned}$$

Then, using the fact that $\|x_1\| \leq \|x_2\|$ we rewrite this as

$$\|\omega_1 x_1 + \omega_2 x_2\|^2 \leq \left(\omega_1^2 + \omega_2^2\right) \|x_2\|^2 + 2\omega_1\omega_2 \langle x_1, x_2 \rangle \tag{24}$$

Applying the Cauchy-Schwarz inequality to the inner product term $\langle x_1, x_2 \rangle$ gives

$$|\langle x_1, x_2 \rangle| \leq \|x_1\| \|x_2\| \leq \|x_2\|^2 \tag{25}$$

with equality if and only if one of x_1 or x_2 is a non-negative scalar multiple of the other and $\|x_1\| = \|x_2\|$, which together imply $x_1 = x_2$ [12].

Multiplying each side of the inequality by $2\omega_1\omega_2$ (which is non-negative since ω_1 and ω_2 are non-negative) gives

$$2\omega_1\omega_2 \langle x_1, x_2 \rangle \leq 2\omega_1\omega_2 \|x_2\|^2$$

Substituting this back into Eqn. (24) to replace the inner product term with its upper bound gives

$$(\omega_1^2 + \omega_2^2)\|x_2\|^2 + 2\omega_1\omega_2 \langle x_1, x_2 \rangle \leq (\omega_1^2 + \omega_2^2)\|x_2\|^2 + 2\omega_1\omega_2 \|x_2\|^2$$

Factoring out $\|x_2\|^2$ on the right-hand side

$$(\omega_1^2 + \omega_2^2)\|x_2\|^2 + 2\omega_1\omega_2 \langle x_1, x_2 \rangle \leq (\omega_1^2 + \omega_2^2 + 2\omega_1\omega_2)\|x_2\|^2$$

Since $\omega_1 + \omega_2 = 1$, we have $\omega_1^2 + \omega_2^2 + 2\omega_1\omega_2 = (\omega_1 + \omega_2)^2 = 1$. Therefore, we can simplify this to

$$(\omega_1^2 + \omega_2^2)\|x_2\|^2 + 2\omega_1\omega_2 \langle x_1, x_2 \rangle = \|\omega_1 x_1 + \omega_2 x_2\|^2 \leq \|x_2\|^2 \tag{26}$$

By Eqn. (25), we have equality in Eqn. (26) if and only if $x_1 = x_2$.

However, we need to generalise this argument to a sum of more than two terms. We proceed by induction on k, the number of terms in the sum. The base case $k = 2$ has already been established in Eqn. (26).

We want to show that for some $k \geq 2$, if the statement in Eqn. (23) holds for $k - 1$ terms, then it also holds for k terms as follows:

Given k vectors $x_1, x_2, \ldots, x_k$ in an inner product space such that $\|x_1\| \leq \|x_2\| \leq \ldots \leq \|x_k\|$, and non-negative weights $\omega_1, \omega_2, \ldots, \omega_k$ such that $\sum_{i=1}^{k} \omega_i = 1$, then

$$\left\| \sum_{i=1}^{k} \omega_i x_i \right\|^2 \leq \|x_k\|^2 \tag{27}$$

with equality if and only if $x_1 = x_2 = \ldots = x_k$.

Assume that for some $k \geq 2$, the inequality holds for $k-1$ terms:

$$\left\| \sum_{i=1}^{k-1} \omega_i x_i \right\|^2 \leq \|x_{k-1}\|^2 \tag{28}$$

with equality if and only if $x_1 = x_2 = \ldots = x_{k-1}$.

Now, we want to show that the inequality also holds for k terms. Starting with the left-hand side of Eqn. (27), we can group the first $k-1$ terms together and the last term separately:

$$\left\| \sum_{i=1}^{k} \omega_i x_i \right\|^2 = \left\| \left(\sum_{i=1}^{k-1} \omega_i x_i \right) + w_k x_k \right\|^2$$

Since the weights are non-negative and sum to 1, we define $W_1 = \sum_{i=1}^{k-1} \omega_i$ and $W_2 = \omega_k$, such that $W_1 + W_2 = 1$, and rewrite the expression as:

$$\left\| \sum_{i=1}^{k} \omega_i x_i \right\|^2 = \left\| W_1 \left(\frac{\sum_{i=1}^{k-1} \omega_i x_i}{\sum_{i=1}^{k-1} \omega_i} \right) + W_2 x_k \right\|^2$$

We define two new vectors y_1 and y_2 as follows:

$$y_1 = \frac{\sum_{i=1}^{k-1} \omega_i x_i}{\sum_{i=1}^{k-1} \omega_i}$$
$$y_2 = x_k$$

Then, our expression becomes

$$\left\| \sum_{i=1}^{k} \omega_i x_i \right\|^2 = \|W_1 y_1 + W_2 y_2\|^2$$

By the inductive hypothesis, we know that $\|y_1\|^2 \leq \|x_{k-1}\|^2$. Since we are given that $\|x_1\| \leq \|x_2\| \leq \ldots \leq \|x_k\|$, it follows that $\|y_1\| \leq \|x_k\| = \|y_2\|$.

Now, we can apply the result of the base case to the vectors y_1 and y_2 with weights W_1 and W_2. By the base case, we have

$$\|W_1 y_1 + W_2 y_2\|^2 \leq \|y_2\|^2 = \|x_k\|^2$$

showing that the inequality in Eqn. (27) holds with equality if and only if $y_1 = y_2$, which implies $x_1 = x_2 = \ldots = x_k$ and completes the induction.

Connecting this back to our original problem, we have shown that the convex combination of vectors $Tx = \sum_{i=1}^{m} \frac{\omega_i}{\omega} S_i x$ satisfies $\|Tx\| \leq \|x\|$ with equality if and only if all the vectors $S_i x$ are equal, which implies that $x \in \mathcal{N}(A)$ (Lemma 1 (i)).

To show that $\|T\| = 1$, we use the definition of the operator norm. The operator norm of T for any non-zero vector x is defined as [10]:

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}$$

This tells us the maximum value by which T can stretch a non-zero vector x [10].

From the inequality $\|Tx\| \leq \|x\|$, dividing both sides by $\|x\|$ (for $x \neq 0$) gives:

$$\frac{\|Tx\|}{\|x\|} \leq 1$$

Taking the supremum over all non-zero x yields:

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} \leq 1 \tag{29}$$

So the operator norm $\|T\|$ is at most 1. However, we want to show that $\|T\| = 1$.

For this, we take an arbitrary non-zero vector $x_0 \in \mathcal{N}(A)$ [13], where we know T leaves x_0 unchanged. Then by Lemma 1 (i), $T(x_0) = x_0$. Therefore, $\|Tx_0\| = \|x_0\|$ and for this specific vector x_0, we have:

$$\frac{\|Tx_0\|}{\|x_0\|} = \frac{\|x_0\|}{\|x_0\|} = 1$$

Therefore, the operator norm $\|T\|$ must be at least 1, i.e. $\|T\| \geq 1$. Combining both results, we have $\|T\| \leq 1$ and $\|T\| \geq 1$, which implies that $\|T\| = 1$. Therefore, T does not expand any vector in $\mathbb{R}^n$ by more than a factor of 1.

Definition 4.3 Non-expansive Mapping

A mapping $T : X \to X$ on a normed vector space X is called non-expansive if for all $x, y \in X$, the following condition holds [14]:

$$\|Tx - Ty\| \leq \|x - y\|$$

Theorem 4.4 Non-expansive Mapping

A linear operator $T : X \to X$ on a normed vector space X is non-expansive if and only if its operator norm satisfies $\|T\| \leq 1$ [14].

($\Rightarrow$) Assume that T is non-expansive, i.e. $\|Tx - Ty\| \leq \|x - y\|$ for all $x, y \in X$. Since $0 \in X$, let $y = 0$, then

$$\|Tx - T0\| \leq \|x - 0\| \implies \|Tx\| \leq \|x\|$$

By Eqn. (29) and the definition of the operator norm, we have $\|T\| \leq 1$.

(⇐) Assume that $\|T\| \leq 1$. Let $z = x - y$ for any $x, y \in X$. Then,

$$\|Tz\| \leq \|z\| \implies \|T(x-y)\| \leq \|x-y\| \implies \|Tx - Ty\| \leq \|x-y\|$$

Therefore, T is a non-expansive mapping, $T : \mathbb{R}^n \to \mathbb{R}^n$ with a maximum stretching factor of 1.

(iii) We want to show that the restriction of T to the subspace $\mathcal{R}(A^T)$, denoted $T\mid_{\mathcal{R}(A^T)}$, is a contraction when $\text{rank}(A) \geq 2$.

Definition 4.5 Restriction of a linear operator

Let $T : V \to W$ be a linear operator and let U be a T-invariant subspace of V. The restriction of T to U, is written as $T|_U : U \to U$ [3, Ch. 9].

From Lemma 1 (ii), we know that $\mathcal{R}(A^T)$ is an invariant subspace of T. Therefore, we can consider the restriction of T to $\mathcal{R}(A^T)$.

From our proof that $\mathcal{N}(T)^\perp = \mathcal{R}(T^*)$ in Theorem 3.11, specifically Eqn. (17), we can uniquely represent the vector space $\mathbb{R}^n$ as the direct sum of $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$:

$$\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A^T) \tag{30}$$

This means that any vector $x \in \mathbb{R}^n$ can be uniquely expressed as $x = n + r$ where $n \in \mathcal{N}(A)$ and $r \in \mathcal{R}(A^T)$. Since $\mathcal{N}(A) \cap \mathcal{R}(A^T) = \{0\}$, then $r \notin \mathcal{N}(A)$ for any non-zero $r \in \mathcal{R}(A^T)$. From Lemma 2 (ii), we have $\|Tx\| \leq \|x\|$ for any $x \in \mathbb{R}^n$, with equality if and only if all the $S_i x$ are equal, implying $x \in \mathcal{N}(A)$.

However, let us consider the implications of the rank condition. If $\text{rank}(A) \geq 2$, then there are at least two linearly independent rows in A. This means that the hyperplanes defined by the equations $A_i x = b_i$ for $i = 1, \ldots, m$ are not all parallel or coincident. So, suppose we have

$$S_1 x = S_i x \quad \forall i = 2, \ldots, m. \tag{31}$$

This would imply that the reflections of x across each hyperplane are the same. However, since we have at least two linearly independent rows in A, this cannot happen unless x is in the null space of A, where T preserves its norm.

By the definition of S_i [2], Eqn. (31) becomes:

$$\left(I - 2\frac{A_1 A_1^T}{\|A_1\|_2^2}\right) x = \left(I - 2\frac{A_i A_i^T}{\|A_i\|_2^2}\right) x$$

Rearranging gives:

$$\frac{A_1^T x}{\|A_1\|_2^2} A_1 - \frac{A_i^T x}{\|A_i\|_2^2} A_i = 0 \tag{32}$$

Since $\text{rank}(A) \geq 2$, then there exists at least one i such that A_i is linearly independent of A_1. The only way for Eqn. (32) to hold is if $A_1^T x = 0$ and $A_i^T x = 0$ [2]. This implies that $Ax = 0$ and thus, $x \in \mathcal{N}(A)$. Therefore, for any non-zero vector $x \in \mathcal{R}(A^T)$, the strict inequality $\|Tx\| < \|x\|$ holds. Hence, T acts as a contraction on the subspace

$\mathcal{R}(A^T)$ when $\text{rank}(A) \geq 2$, and we can conclude that $\|T|_{\mathcal{R}(A^T)}\| < 1$ by the definition of the operator norm.

If $\text{rank}(A) = 1$, all the rows of A are scalar multiples of each other, meaning the hyperplanes defined by each row equation have parallel normals A_i [3, Ch. 4]. Consequently, the reflection operator S_i for each row will be identical, and T defines a convex combination of identical operators S_i. Therefore, $\|Tx\| = \|x\|$ for any $x \in \mathbb{R}^n$. This means that T does not act as a contraction on $\mathcal{R}(A^T)$ when $\text{rank}(A) = 1$, and hence, we cannot guarantee that $\|T|_{\mathcal{R}(A^T)}\| < 1$.

Geometrically, when $\text{rank}(A) = 1$, we will see an approximation x oscillating between two points on the same hypersphere [2], as applying the same reflection twice results in the original point, i.e. $S_1(S_1x) = x$. In this case, the norm of the component of x in $\mathcal{R}(A^T)$ is not reduced since $\mathcal{R}(A^T)$ is one-dimensional and spanned by the single row vector [15].

This becomes very clear in two-dimensional space. Consider the case where the system is defined by a single row vector $A_1 = [1, 1]$ and $b_1 = -2$. The solution space is the hyperplane $x_1 + x_2 = -2$ and is a translation of $\mathcal{N}(A_1)$ by a particular solution to $A_1x = b_1$ [16]. We take an arbitrary approximation $x = [0, 0]$.

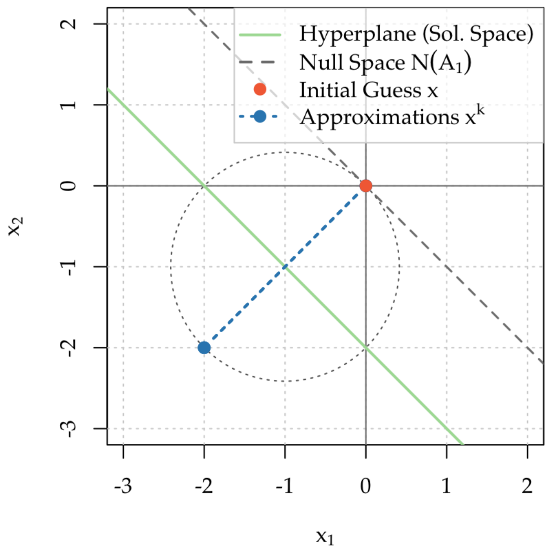


Figure 2: Cimmino's algorithm for $A_1 = [1, 1]$, $b_1 = -2$, and solution space $x_1 + x_2 = -2$. The point $x = [0, 0]$ oscillates between two points on a hypersphere.

(iv) Finally, we want to show that $\|T(x)\| = \|x\|$ if and only if $x \in \mathcal{N}(A)$.

This follows directly from the discussion in Lemma 2 (ii): $\|Tx\| \leq \|x\|$ for any $x \in \mathbb{R}^n$, with equality if and only if all the S_ix are equal, implying $x \in \mathcal{N}(A)$.

Therefore, if $\|Tx\| = \|x\|$, it must be that $x \in \mathcal{N}(A)$. Conversely, if $x \in \mathcal{N}(A)$, then from Lemma 1 (i), we know that $T(x) = x$ and therefore, $\|Tx\| = \|x\|$.

We have established that T is a non-expansive operator in $\mathbb{R}^n$ that acts as the identity on $\mathcal{N}(A)$ and as a contraction on $\mathcal{R}(A^T)$ when $\text{rank}(A) \geq 2$. We now extend our analysis to the recursive application of T over k iterations.

5 Recursive Application of T

In this section, we derive a closed-form expression for T^k [2], where k is a positive integer, and decompose the effect of T on any vector $x \in \mathbb{R}^n$ into two components: one that remains unchanged in the null space of A, and another that is progressively contracted within the range of A^T. This decomposition provides insight into the iterative behaviour of Cimmino's algorithm and its convergence properties.

Lemma 6. *Given the linear operator T as defined in Eqn. (10), the following recursive relationship holds for any integer $k \geq 1$:*

$$T^k = P_{\mathcal{N}(A)} + \widetilde{T}^k \tag{33}$$

where $\widetilde{T} = TP_{\mathcal{R}(A^T)}$ and $T^k = TT^{k-1}$ with $T^0 = I$ [2].

We proceed with an inductive proof to establish the recursive relationship in Lemma 6. First, we verify the base case when $k = 1$ [2]:

$$T^1 = TI = T(P_{\mathcal{N}(A)} + P_{\mathcal{R}(A^T)}) = TP_{\mathcal{N}(A)} + TP_{\mathcal{R}(A^T)} \overset{\text{Lemma 3 }(i)}{=} P_{\mathcal{N}(A)} + \widetilde{T} \tag{34}$$

Assume that Lemma 6 holds for some integer $k \geq 1$. We want to show that if Eqn. (33) holds for k, then it also holds for $k + 1$ [2].

$$\begin{aligned} T^{k+1} = TT^k &= \left(P_{\mathcal{N}(A)} + \widetilde{T}\right)\left(P_{\mathcal{N}(A)} + \widetilde{T}^k\right) \\ &= P^2_{\mathcal{N}(A)} + P_{\mathcal{N}(A)}\widetilde{T}^k + \widetilde{T}P_{\mathcal{N}(A)} + \widetilde{T}^{k+1} \end{aligned}$$

We can substitute the definition of $\widetilde{T} = TP_{\mathcal{R}(A^T)}$ into $\widetilde{T}P_{\mathcal{N}(A)}$:

$$T^{k+1} = P^2_{\mathcal{N}(A)} + P_{\mathcal{N}(A)}\widetilde{T}^k + TP_{\mathcal{R}(A^T)}P_{\mathcal{N}(A)} + \widetilde{T}^{k+1}$$

From our earlier results in Lemma 3 (i), we know that T acts as the identity on $\mathcal{N}(A)$, thus $TP_{\mathcal{N}(A)} = P_{\mathcal{N}(A)}$. By Theorem 3.9, we have $P_{\mathcal{N}(A)}P_{\mathcal{R}(A^T)} = 0$.

By a similar argument, we can show that $P_{\mathcal{N}(A)}\widetilde{T}^k = 0$. Isolating the term $P_{\mathcal{N}(A)}\widetilde{T}$ and applying it to an arbitrary vector $x \in \mathbb{R}^n$ gives [2]:

$$P_{\mathcal{N}(A)}\widetilde{T}x = P_{\mathcal{N}(A)}TP_{\mathcal{R}(A^T)}x$$

The projection of x onto $\mathcal{R}(A^T)$, given by $P_{\mathcal{R}(A^T)}x$, is a vector in $\mathcal{R}(A^T)$. By Lemma 3 (ii), we know that T maps vectors in $\mathcal{R}(A^T)$ back into $\mathcal{R}(A^T)$. Therefore, $TP_{\mathcal{R}(A^T)}x \in \mathcal{R}(A^T)$, and it follows that $P_{\mathcal{N}(A)}TP_{\mathcal{R}(A^T)}x = 0$.

Therefore, $P_{\mathcal{N}(A)}\widetilde{T} = 0$, and by extension, $P_{\mathcal{N}(A)}\widetilde{T}^k = 0$ for any integer $k \geq 1$. As a result, we have:

$$T^{k+1} = P^2_{\mathcal{N}(A)} + 0 + 0 + \widetilde{T}^{k+1}$$

Since $P_{\mathcal{N}(A)}$ is a projection operator, it is idempotent, meaning that $P^2_{\mathcal{N}(A)} = P_{\mathcal{N}(A)}$ [17]. Therefore, we can simplify the expression to:

$$T^{k+1} = P_{\mathcal{N}(A)} + \widetilde{T}^{k+1}$$

This completes the inductive step, showing that if Lemma 6 holds for k, then it also holds for $k+1$ and thus, holds for all integers $k \geq 1$. What we have established is the long-term behaviour of the operator T when applied recursively. We have two key observations from this result: for any vector $x \in \mathbb{R}^n$, the norm of the component that lies in the null space of A, $P_{\mathcal{N}(A)}x$, is preserved, while the component in the range of A^T, $P_{\mathcal{R}(A^T)}x$ is transformed by $\widetilde{T}$ at each iteration, strictly reducing its norm.

With Lemma 6 verified, we proceed to the final part of our analysis, where we formally establish the convergence of Cimmino's algorithm and characterise its limit point.

6 Convergence and Limit of Cimmino's Algorithm

In the following section, we present and prove the main theorem regarding the convergence of Cimmino's algorithm and the characterisation of its limit point. This theorem encapsulates the results we have derived so far and provides a comprehensive understanding of the algorithm's behaviour under the specified conditions.

Theorem 6.1 Convergence and Limit of Cimmino's Algorithm

Given a matrix A that satisfies the conditions in Eqn. (1) then, for any initial approximation $x^0 \in \mathbb{R}^n$, Cimmino's algorithm produces a convergent sequence of approximations $\{x^k\}$ and its limit as k approaches infinity is given by [2]:

$$\lim_{k\to\infty} x^k = P_{\mathcal{N}(A)}(x^0) + \left(I - \widetilde{T}\right)^{-1} Rb \tag{35}$$

If the system is consistent, then,

$$\left(I - \widetilde{T}\right)^{-1} Rb = x_{LS} \tag{36}$$

and the limit point in Eqn. (35) is one of the solutions to the system [2].

Theorem 6.1 states that the constructed sequence of approximations $\{x^k\}$ converges to a specific limit point, which is the sum of two components: the projection of the initial approximation onto the null space of A and the term $(I - \widetilde{T})^{-1}Rb$, which in a consistent system is the minimum norm solution x_{LS}. In other words, Cimmino's algorithm will always converge to a solution of $Ax = b$ if the system is consistent, and if not, it will converge to a point that minimises the weighted residual $\|D(Ax - b)\|_2$.

We start by recursively expanding x^k using the expression, $x^{k+1} = Tx^k + Rb$ from Eqn. (9) [2], to express x^k in terms of the initial approximation x^0 and the cumulative effect of applying the operator T to the term Rb:

$$\begin{aligned} x^k &= Tx^{k-1} + Rb \\ &= T(Tx^{k-2} + Rb) + Rb \\ &= T(Tx^{k-3} + Rb) + TRb + Rb \\ &\vdots \\ &= T^k x^0 + \sum_{j=0}^{k-1} T^j Rb \end{aligned} \tag{37}$$

In other words, we continuously substitute the expression for x^k back into itself, each time replacing x^{k-1} with $Tx^{k-2} + Rb$, and so on, until we reach the initial approximation x^0. This process effectively unravels the recursive definition of x^k, resulting in a sum of the effects of applying T k times to x^0 and the accumulated contributions from the term Rb at each iteration.

From Lemma 6, we know that $T^k = P_{\mathcal{N}(A)} + \widetilde{T}^k$. Substituting this into Eqn. (37) gives [2]:

$$\begin{aligned} x^k &= \left(P_{\mathcal{N}(A)} + \widetilde{T}^k\right) x^0 + \sum_{j=0}^{k-1} \left(P_{\mathcal{N}(A)} + \widetilde{T}^j\right) Rb \\ &= P_{\mathcal{N}(A)} x^0 + \widetilde{T}^k x^0 + \sum_{j=0}^{k-1} P_{\mathcal{N}(A)} Rb + \sum_{j=0}^{k-1} \widetilde{T}^j Rb \end{aligned}$$

By Lemma 3 (iii), $Rb \in \mathcal{R}(A^T)$, and since $\mathcal{N}(A) \perp \mathcal{R}(A^T)$ (Theorem 3.9), we have $P_{\mathcal{N}(A)} Rb = 0$. Therefore, simplifying our expression to:

$$x^k = P_{\mathcal{N}(A)} x^0 + \widetilde{T}^k x^0 + \sum_{j=0}^{k-1} \widetilde{T}^j Rb \tag{38}$$

Taking a closer look at the term $\widetilde{T} = TP_{\mathcal{R}(A^T)}$, we note that it is a composition of two operators: the orthogonal projector $P_{\mathcal{R}(A^T)}$ and the operator T.

For any vector $x \in \mathbb{R}^n$, we consider the action of $P_{\mathcal{R}(A^T)}$ on x.

Let P be an orthogonal projector onto a subspace V of $\mathbb{R}^n$. For any vector $x \in \mathbb{R}^n$, Px represents the projection of x onto V. By Theorem 3.9, we have the decomposition of x into its orthogonal components [3, Ch. 8]:

$$x = (x - Px) + Px$$

where $x - Px \in V^\perp$ and $Px \in V$.

As a result of the orthogonality of V and $V^\perp$, we can directly apply the Pythagorean theorem to relate the norms of these components [3, Ch. 5, Ch. 8]:

$$\begin{aligned} \|x\|^2 &= \|Px\|^2 + \|x - Px\|^2 \\ \|Px\|^2 &= \|x\|^2 - \|x - Px\|^2 \end{aligned}$$

Since $\|x - Px\|^2$ is non-negative [3, Ch. 10], we have:

$$\|Px\|^2 \leq \|x\|^2$$
$$\|Px\| \leq \|x\|$$

Taking the supremum over all non-zero x gives

$$\|P\| = \sup_{x \neq 0} \frac{\|Px\|}{\|x\|} \leq 1 \tag{39}$$

for any orthogonal projector P onto a subspace of $\mathbb{R}^n$. Therefore, $\|P_{\mathcal{R}(A^T)}\| \leq 1$.

From Lemma 4 (iii), we know that T is a contraction mapping on $\mathcal{R}(A^T)$ when $\operatorname{rank}(A) \geq 2$ and by definition $P_{\mathcal{R}(A^T)}$ maps any vector in $\mathbb{R}^n$ onto $\mathcal{R}(A^T)$. Consequently,

$$\|\widetilde{T}\| = \|TP_{\mathcal{R}(A^T)}\| < 1 \tag{40}$$

establishing that $\widetilde{T}$ is also a contraction mapping on $\mathcal{R}(A^T)$.

Now, going back to Eqn. (38), we can observe that as k approaches infinity, the term $\widetilde{T}^k x^0$ will converge to zero because $\|\widetilde{T}\| < 1$. Hence, we have:

$$\lim_{k \to \infty} \widetilde{T}^k x^0 = 0 \tag{41}$$

Next, we consider the partial sum $\sum_{j=0}^{k-1} \widetilde{T}^j Rb$ of the Neumann series $\sum_{j=0}^{\infty} \widetilde{T}^j$ applied to the vector Rb [2], [18].

Definition 6.2 Neumann Series

Let T be a bounded linear operator on a normed vector space W such that $\|T\| < 1$ [18]. The Neumann series is defined as the infinite series:

$$\sum_{j=0}^{\infty} T^j = I + T + T^2 + T^3 + \dots$$

where I is the identity operator on W. The Neumann series converges in the operator norm and its sum is given by [18]:

$$(I - T)^{-1} = \sum_{j=0}^{\infty} T^j$$

We have already established that $\|\widetilde{T}\| < 1$ in Eqn. (40) hence, $I - \widetilde{T}$ is invertible and the Neumann series $\sum_{j=0}^{\infty} \widetilde{T}^j$ converges [2], [18]. Therefore, by Definition 6.2, we obtain:

$$\lim_{k \to \infty} \sum_{j=0}^{k-1} \widetilde{T}^j Rb = (I - \widetilde{T})^{-1} Rb \tag{42}$$

Combining Eqn. (41) and Eqn. (42) and taking the limit of x^k in Eqn. (38) as k approaches infinity:

$$\begin{aligned}\lim_{k\to\infty} x^k &= \lim_{k\to\infty}\left(P_{\mathcal{N}(A)}x^0 + \widetilde{T}^k x^0 + \sum_{j=0}^{k-1}\widetilde{T}^j Rb\right)\\ &= P_{\mathcal{N}(A)}x^0 + 0 + (I-\widetilde{T})^{-1}Rb\\ &= P_{\mathcal{N}(A)}x^0 + (I-\widetilde{T})^{-1}Rb\end{aligned}$$

These results provide the exact limit of the sequence $\{x^k\}$ produced by Cimmino's algorithm, as described in Eqn. (35) of Theorem 6.1 [2].

Finally, we consider the specific case when the system of equations $Ax = b$ is consistent, i.e. $b \in \mathcal{R}(A)$. In this case, Theorem 6.1 states that Cimmino's algorithm converges to a solution of the system $P_{\mathcal{N}(A)}x^0 + x_{LS}$, where $x_{LS} = \left(I - \widetilde{T}\right)^{-1} Rb$ is the minimum norm solution such that $\|x_{LS}\|_2 \leq \|x\|_2$ for all solutions x [19].

We want to show that in a consistent system, the term $(I-\widetilde{T})^{-1}Rb$ is equivalent to x_{LS} as stated in Eqn. (36). To do this, we will use the generalised inverse of a matrix [2].

Definition 6.3 Generalised Inverse

If A is a real $m \times n$ matrix, then the generalised inverse of A, denoted by G, is an $n \times m$ matrix satisfying the condition [19]:

$$AGA = A \tag{43}$$

Our goal is to show that $x^* = (I-\widetilde{T})^{-1}Rb$ is the minimum norm solution, x_{LS}, to the consistent system $Ax = b$, i.e.

$$(I-\widetilde{T})^{-1}Rb = x_{LS} \tag{44}$$

To achieve this, we will demonstrate two properties of x^*:

(i) x^* is a solution to the system $Ax = b$, and

(ii) x^* lies in the range of A^T, i.e. $x^* \in \mathcal{R}(A^T)$ [19].

(i) To prove that x^* is a solution to the system $Ax = b$, we need to show that $Ax^* = b$.

We define the generalised inverse G of A as [2]:

$$G := (I-\widetilde{T})^{-1}R \tag{45}$$

and write $x^* = Gb$.

Since the system $Ax = b$ is assumed to be consistent, then $b \in \mathcal{R}(A)$ and equivalently $AGb = b$, where G is the generalised inverse of A satisfying $AGA = A$ [2], [19]. We want to show that $G = (I-\widetilde{T})^{-1}R$ is indeed a generalised inverse of A satisfying the condition $AGA = A$ and thus, $Gb = x^*$ is a solution to $Ax = b$ [20].

We first establish that the definitions of the operators T and R satisfy the following expression [2]:

$$I - T = RA \tag{46}$$

Substituting the expression for R from Eqn. (11) into RA:

$$RA = \left(\frac{2}{\omega}A^T D^T D\right) A = \frac{2}{\omega}A^T D^T DA$$

Substituting the expression for T from Eqn. (10) into $I - T$:

$$I - T = I - I + \frac{2}{\omega}A^T D^T DA = \frac{2}{\omega}A^T D^T DA$$

Therefore, we have $I - T = RA$.

Next, we compute AGA using Eqn. (45) and Eqn. (46) [2]:

$$\begin{aligned}
AGA &= A(I - \widetilde{T})^{-1}RA \\
&= A(I - \widetilde{T})^{-1}(I - T) \quad \text{(Eqn. (46))} \\
&= A(I - \widetilde{T})^{-1}((I - \widetilde{T}) - P_{\mathcal{N}(A)}) \quad \text{(Lemma 6)} \\
&= A(I - \widetilde{T})^{-1}(I - \widetilde{T}) - A(I - \widetilde{T})^{-1}P_{\mathcal{N}(A)} \\
&= A - A(I - \widetilde{T})^{-1}P_{\mathcal{N}(A)} \\
&= A - A\sum_{j=0}^{\infty}\widetilde{T}^j P_{\mathcal{N}(A)} \quad \text{(Definition 6.2)} \\
&= A - 0 \quad \text{(Definition 3.2 and } \widetilde{T}\text{)} \\
&= A
\end{aligned}$$

We have shown that $AGA = A$ and can conclude that $G = (I - \widetilde{T})^{-1}R$ is a generalised inverse of A. It directly follows that $Gb = x^* = (I - \widetilde{T})^{-1}Rb$ is a solution to the consistent system $Ax = b$ [2], [19], thus proving property (i).

(ii) To show that $x^* \in \mathcal{R}(A^T)$, we draw from our earlier results in Lemma 3. By Lemma 3 (iii), $Rb \in \mathcal{R}(A^T)$. By Lemma 3 (ii) and the definition of $\widetilde{T} = TP_{\mathcal{R}(A^T)}$, it follows that $\mathcal{R}(A^T)$ is invariant under $\widetilde{T}$. Consequently, $(I - \widetilde{T})^{-1}$ also maps $\mathcal{R}(A^T)$ into itself. Therefore, the composition $(I - \widetilde{T})^{-1}Rb$ lies entirely within $\mathcal{R}(A^T)$, establishing that $x^* \in \mathcal{R}(A^T)$ and thus proving property (ii).

From the decomposition $\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A^T)$ (Theorem 3.9 and 3.11), it follows that any vector in $\mathcal{R}(A^T)$ is orthogonal to $\mathcal{N}(A)$. Since $x^* \in \mathcal{R}(A^T)$ and is orthogonal to $\mathcal{N}(A)$, it is the minimum norm solution, x_{LS}, to the consistent system $Ax = b$ [20]. We have shown that for a consistent system, $(I - \widetilde{T})^{-1}Rb = x_{LS}$, therefore proving Eqn. (36).

As a result, the limit point of the sequence $\{x^k\}$ in Eqn. (35) is given by $P_{\mathcal{N}(A)}(x^0) + x_{LS}$, which is the sum of a vector in the null space of A and the minimum norm solution x_{LS}. This is precisely the general form of a solution in the consistent case where any solution is given by the sum of a particular solution and a vector from the null space of A [16], [20]. This result directly satisfies Theorem 1.1 (i) and we conclude that the sequence $\{x^k\}$ produced by Cimmino's algorithm converges to a solution of the system when it is consistent [2].

It follows that the limit point of the sequence $\{x^k\}$ for an inconsistent system is given by $P_{\mathcal{N}(A)}(x^0) + x_{LS}$, where $x_{LS} = (I - \widetilde{T})^{-1}Rb$ minimises the weighted residual norm $\|D(Ax - b)\|_2$. This result addresses Theorem 1.1 (ii) and we conclude that the generated sequence $\{x^k\}$ in Cimmino's algorithm converges to a point that minimises the weighted residual norm when the system is inconsistent [2].

This completes the proof of Theorem 6.1 and our analysis on the convergence and limit of Cimmino's algorithm.

7 Application to Image Reconstruction

Cimmino's algorithm has practical applications in various fields, including image reconstruction, where it is used to solve large-scale linear systems that arise from tomographic imaging techniques. In this section, we will illustrate the application of Cimmino's algorithm to a well-known image reconstruction problem, the Shepp-Logan phantom, which is a standard image used for Computed Tomography (CT) simulations [21]. Our intention is not to provide a comprehensive overview of image reconstruction techniques or achieve cutting-edge results, but rather to demonstrate the effectiveness of Cimmino's algorithm in solving the associated linear systems.

The Shepp-Logan phantom is a synthetic image that simulates the cross-sectional view of a human head, comprising of ten ellipses of varying intensities and sizes [21]. The goal of image reconstruction is to recover the original image from its projections, which are obtained by simulating the passage of X-rays through the phantom at different angles [22].

The model for the image reconstruction problem can be formulated as a linear system $Ax = b$, where:

> ***A*:** System matrix that represents the projection operation, mapping the image space to the projection space [23]. M is the total number of rays given by the number of projection views (angles) multiplied by the number of detectors, and N is the total number of pixels in the image.
>
> ***x*:** Unknown image vector (column vector of pixel intensities) of length N.
>
> ***b*:** Vector of measured projections, also known as the sinogram, of length M.

The main challenge with such image reconstruction problems is that the system matrix A is often very large and sparse and the system $Ax = b$ is typically inconsistent due to noise in the measurements or incomplete data [21]. Direct methods for solving such systems can be computationally expensive and impractical, making iterative methods like Cimmino's algorithm more suitable [2].

For the following image reconstruction problem, we use the following parameters:

> **Image size:** 512×512, resulting in $N = 262144$ pixels.

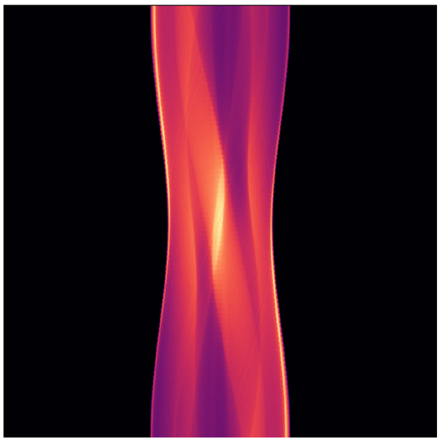
Number of angles: 90 evenly spaced angles between 0 and 180 degrees.

Number of detectors: $\lceil 2\sqrt{2} * 512 \rceil = 1449$ [22].

The system matrix A is constructed in Python using the ASTRA-Toolbox library [24]. The matrix A is generated using a 2D parallel beam geometry with a 'strip' projector type [25] and is stored as a sparse matrix. The Shepp-Logan phantom is also generated using the ASTRA-Toolbox library.

The sinogram b is obtained by projecting the Shepp-Logan phantom using the system matrix A, i.e. $b = AP$, where P is the vectorised Shepp-Logan phantom. Figures 3a and 3b show the Shepp-Logan phantom and its associated sinogram, respectively.

(a) Shepp-Logan Phantom (512×512) generated using ASTRA-Toolbox [24].



(b) Sinogram b (projections) of the Shepp-Logan Phantom P. Computed as $b = AP$.

Figure 3: Visualisation of the Shepp-Logan Phantom alongside its associated sinogram.

Our implementation of Cimmino's algorithm is executed in C++/Metal to leverage GPU parallelisation as our problem size is large, with the system matrix A containing approximately 75 million non-zero elements.

The implementation follows the algorithmic steps outlined in Section 2, with the following specific choices:

Weights: $\omega_i = \|A_i\|_2^2$ for $i = 1, 2, \ldots, M$, as in Eqn. (7), and $\omega = \sum_{i=1}^{M} \omega_i$.

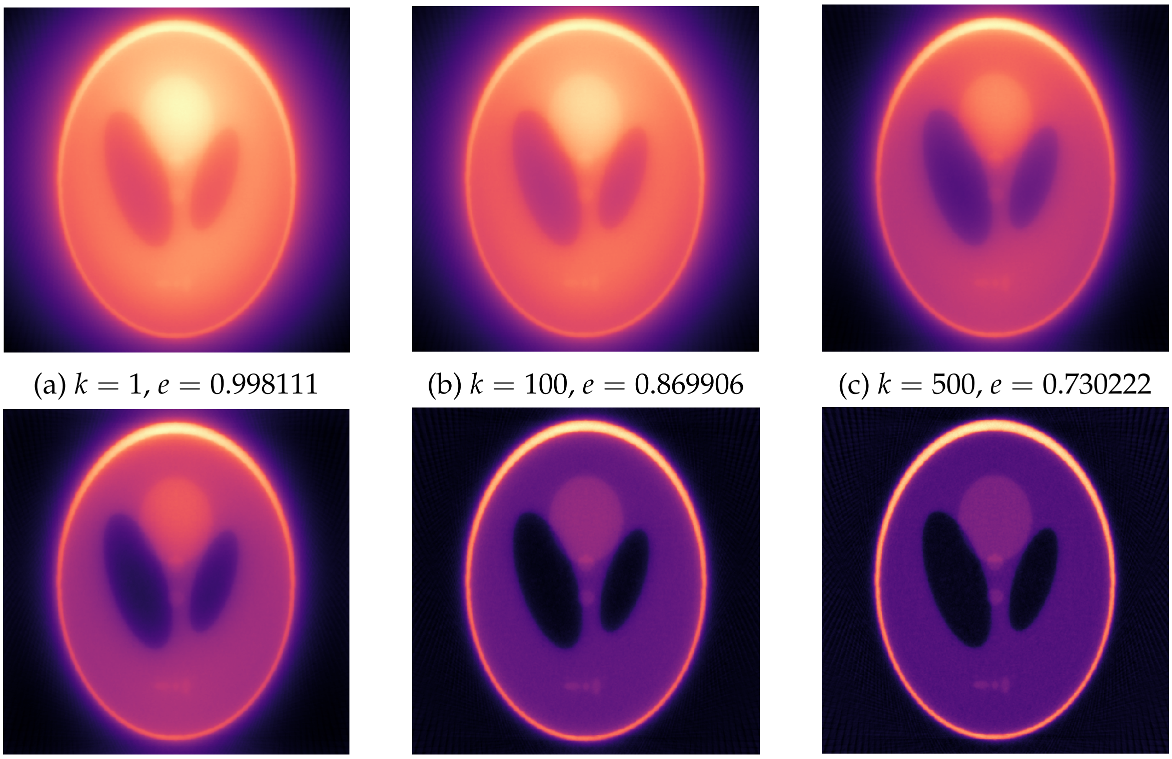
Initial approximation: x^0 is chosen as a zero vector of length N (all pixel intensities set to zero).

Stopping criterion: The algorithm is run until the maximum number of iterations k is reached or the relative error norm [2], [26] between the original phantom P and the reconstructed image x^k is less than a tolerance of 10^{-2}, i.e.

$$e = \frac{\|x^k - P\|_2}{\|P\|_2} < 10^{-2}$$

where P is the vectorised Shepp-Logan phantom.

We obtain the reconstructed image x^k for a fixed number of iterations, $k = 1, 100, 500, 1000, 5000$ and 10000 to illustrate the convergent nature of the algorithm. The relative error norm e is also computed at each of these iterations to quantify the improvement in the reconstruction over different k. The results are shown in Figure 4.



(a) $k = 1$, $e = 0.998111$ (b) $k = 100$, $e = 0.869906$ (c) $k = 500$, $e = 0.730222$

(d) $k = 1000$, $e = 0.660265$ (e) $k = 5000$, $e = 0.436225$ (f) $k = 10000$, $e = 0.328996$

Figure 4: Reconstructed Shepp-Logan phantoms using Cimmino's algorithm in C++/Metal for $k = 1, 100, 500, 1000, 5000, 10000$ iterations with relative error norms e.

As the number of iterations k increases, the reconstructed image x^k progressively approximates the original Shepp-Logan phantom P, demonstrating the convergent nature of Cimmino's algorithm. The relative error norm e decreases with increasing k, indicating that the reconstruction quality improves as the number of iterations increases. The results in Figure 4 are modest and achieving more refined reconstructions would likely require a greater number of angles, detectors, and iterations, as well as finetuning parameters, preconditioning techniques and optimisation strategies [21], [23].

Our demonstration of Cimmino's algorithm for the Shepp-Logan image reconstruction problem highlights only one of its many applications. The algorithm's convergent properties and ability to handle large, sparse, and inconsistent systems make it a valuable tool in various scientific domains, including medical imaging, digital signal processing, and holography [27].

In this report, we presented a comprehensive analysis of Cimmino's algorithm, emphasising its convergence properties in both consistent and inconsistent linear systems. Our analysis and discussions highlight the algorithm's mathematical elegance and practical utility.

8 Acknowledgments

I want to acknowledge Julien and Lashi for their invaluable guidance and support throughout this project. In particular, for their in-depth feedback and guidance on the proofs throughout Sections 3 and 4, as well as answering my endless questions across all areas of the report. Thank you to Simon for supporting this endeavour!

References

[1] I. A. Ferreira, J. A. Acebrón, and J. Monteiro, *Survey of a Class of Iterative Row-Action Methods: The Kaczmarz Method*, 2024. arXiv: 2401.02842 [math.NA]. [Online]. Available: https://arxiv.org/abs/2401.02842.

[2] S. Petra, C. Popa, and C. Schnörr, "Extended and constrained cimmino-type algorithms with applications in tomographic image reconstruction," Nov. 2008. DOI: 10.11588/heidok.00008798.

[3] W. K. Nicholson, "Linear algebra with applications," in Open edition, Version 2021 - Revision A. Lyryx Learning Inc., Dec. 2020. [Online]. Available: https://lyryx.com/linear-algebra-applications/.

[4] M. Barile, *Affine*, MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. [Online]. Available: https://mathworld.wolfram.com/Affine.html.

[5] nLab authors, *Reflection at a hyperplane*, Revision 2, Aug. 2025. [Online]. Available: https://ncatlab.org/nlab/show/reflection+at+a+hyperplane.

[6] V. Dobrushkin. "Linear algebra with mathematica. part 3: Vector spaces - direct sums." (2008), [Online]. Available: https://www.cfm.brown.edu/people/dobrush/cs52/Mathematica/Part3/direct.html.

[7] T. Miyake, *Linear algebra : form the beginnings to the Jordan normal forms / Toshitsune Miyake.* Springer, 2022. [Online]. Available: https://research.ebsco.com/linkprocessor/plink?id=4244a62e-d4ad-3450-98a5-d8c0d58433bc.

[8] D. Bindel, *Householder transformations*, Cornell CS 6210 Matrix Computations, 2019.

[9] T. Rowland and E. W. Weisstein, *Orthogonal Matrix*, MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. [Online]. Available: https://mathworld.wolfram.com/OrthogonalMatrix.html.

[10] nLab authors, *Operator norm*, Aug. 2025. [Online]. Available: https://ncatlab.org/nlab/show/operator+norm.

[11] R. Hedegaard, *Convex combination*, MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. [Online]. Available: https://mathworld.wolfram.com/ConvexCombination.html.

[12] A. Farahmand, *6.7 cauchy-schwarz inequality*, UC Berkeley. Department of Mathematics. [Online]. Available: https://math.berkeley.edu/~arash/54/notes/.

[13] K. CONRAD, *Computing the norm of a matrix*, Department of Mathematics - University of Connecticut. [Online]. Available: https://kconrad.math.uconn.edu/blurbs/.

[14] H. Bauschke and J. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, 1995. DOI: 10.1137/S0036144593251710.

[15] J. M. Belk, *10.1 subspaces*, Cornell Mathematics Department - Cornell University, 2021. [Online]. Available: https://e.math.cornell.edu/people/belk/writing/.

[16] D. Margalit and J. Rabinoff. "2.4 solution sets." (2019), [Online]. Available: https://textbooks.math.gatech.edu/ila/solution-sets.html.

[17] V. Dobrushkin, *Projection operators*. [Online]. Available: https://www.cfm.brown.edu/people/dobrush/am34/MuPad/projection.html.

[18] Y. Zhang and T. Trogdon, "A probabilistic analysis of the Neumann series iteration," *Minn. J. Undergrad. Math.*, 2022.

[19] G. Chen. "Lecture 6: Generalized inverse and pseudoinverse." (2025), [Online]. Available: https://www.sjsu.edu/faculty/guangliang.chen/.

[20] K. Tanabe, "Projection method for solving a singular system of linear equations and its applications.," *Numer. Math.*, vol. 17, pp. 203–214, 1971. DOI: 10.1007/BF01436376.

[21] H. Gach, C. Tanase, and F. Boada, "2d and 3d shepp-logan phantom standards for mri," in *19th International Conference on Systems Engineering*, Sep. 2008, pp. 521–526. DOI: 10.1109/ICSEng.2008.15.

[22] A. C. Kak and M. Slaney, "Algorithms for reconstruction with nondiffracting sources," in *Principles of Computerized Tomographic Imaging*, ch. 3, pp. 49–112. DOI: 10.1137/1.9780898719277.ch3.

[23] G. Mahmoudi, M. R. Ay, A. Rahmim, and H. Ghadiri, "Computationally efficient system matrix calculation techniques in computed tomography iterative reconstruction.," *J. Med. Signals Sens.*, vol. 10, 2020. DOI: 10.4103/jmss.JMSS_29_19.

[24] "The ASTRA Toolbox." (), [Online]. Available: https://astra-toolbox.com/.

[25] "2D Geometries - Volume Geometries." (), [Online]. Available: https://astra-toolbox.com/docs/geom2d.html#parallel.

[26] Y. Han, H. D. F. Lam, W. Mei, and L. Fang, "Image reconstruction using analysis model prior," *Comput. Math. Method. M.*, 2016. DOI: 10.1155/2016/7571934.

[27] M. de V. El Guendy, R. Mouthaan, A. Kadis, *et al.*, "Cimmino simultaneously iterative holographic projection," *Opt. Continuum*, vol. 1, no. 6, pp. 1351–1363, Jun. 2022. DOI: 10.1364/OPTCON.451113.