Cimmino's method and the next generation of iterative solvers

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

The Problem

Approximation to the solution $\mathbf{x}^*$ of

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

with

⋆ $\mathbf{A} \in \mathbb{R}^{n\times m}$, $n \leq m$

⋆ $\mathbf{b} \in \mathrm{range}(\mathbf{A})$

Given $\mathbf{x}^{(0)}$, generate sequence

$$\{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots\}, \qquad \mathbf{x}^{(k)} \rightarrow \mathbf{x}^*$$

- We start with $n = m$ ($\mathbf{A}$ square)
- The solution of $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ will also be considered

Projection Methods

$$\mathbf{Ax} = \mathbf{b}$$

Choose $\mathcal{K}$ such that

$$\mathbf{x}^{(k)} \in \mathcal{K}, \quad \mathbf{x}^{(k)} \approx \mathbf{x}^*$$

Projection Methods

$$\mathbf{Ax} = \mathbf{b}$$

Choose $\mathcal{K}$ such that

$$\mathbf{x}^{(k)} \in \mathcal{K}, \quad \mathbf{x}^{(k)} \approx \mathbf{x}^*$$

Various alternatives for $\mathcal{K}$:

- Generate sequence of $\mathcal{K}_k \subset \mathcal{K}_{k+1}$ and impose a global optimality condition. E.g.

  $$\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{Ax}^{(k)} \perp \mathcal{K}_k, \quad k = 1, 2, \ldots$$

  (Krylov subspace methods...)

-

Projection Methods

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

Choose $\mathcal{K}$ such that

$$\mathbf{x}^{(k)} \in \mathcal{K}, \quad \mathbf{x}^{(k)} \approx \mathbf{x}^*$$

Various alternatives for $\mathcal{K}$:

- Generate sequence of $\mathcal{K}_k \subset \mathcal{K}_{k+1}$ and impose a global optimality condition. E.g.

  $$\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)} \perp \mathcal{K}_k, \quad k = 1, 2, \ldots$$

  (Krylov subspace methods...)

- Fix large space $\mathcal{K}$ with $\mathbf{x}^* \in \mathcal{K}$ and select sequence of $\mathbf{x}^{(k)}$ satisfying a local optimality condition.

  (Stationary iterative methods...)

Geometric derivation. I

A simplified case. $n = 2$

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \begin{cases} a_{1,1}x_1 + a_{1,2}x_2 = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 = b_2 \end{cases}$$

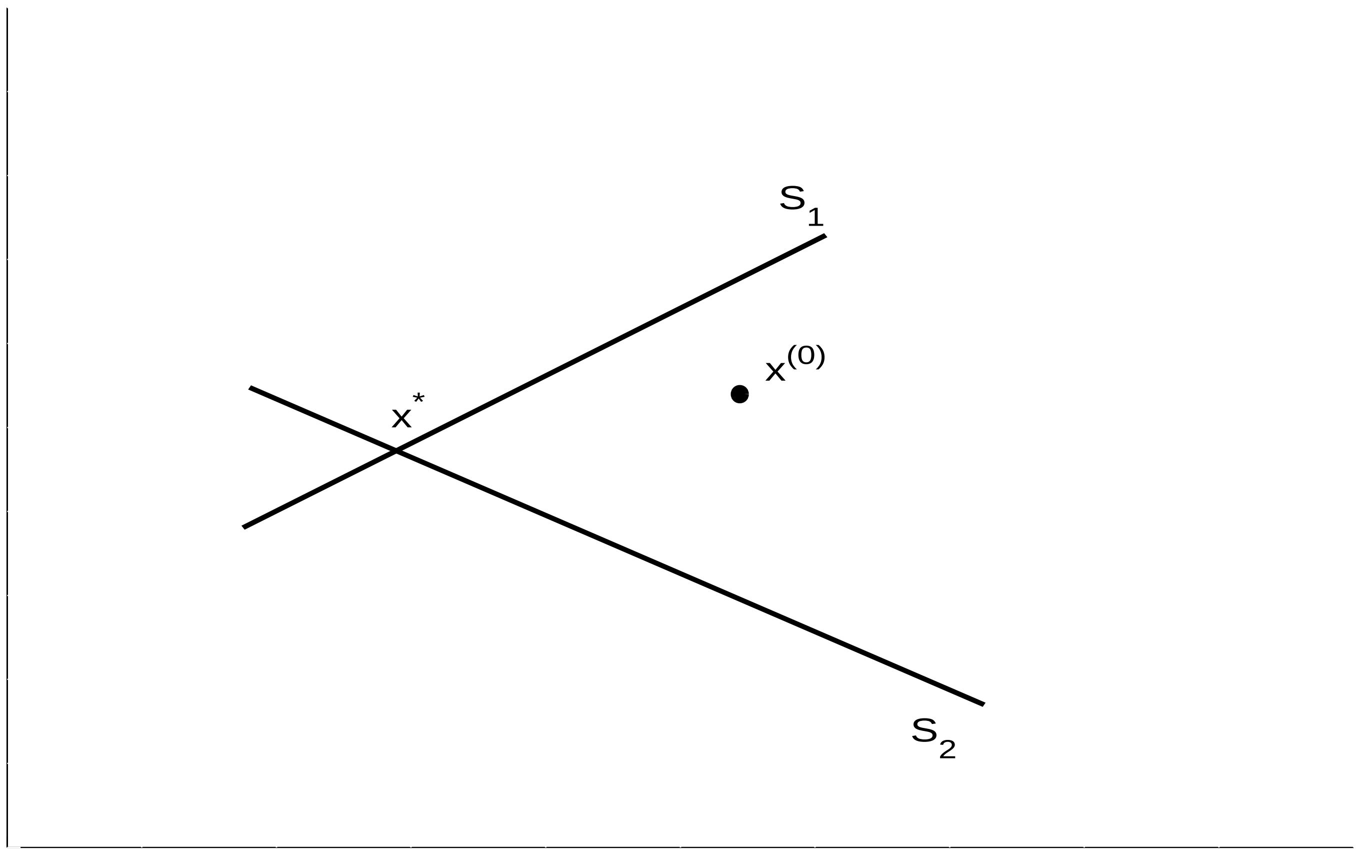
$\mathcal{S}_1 = \{\mathbf{x} \in \mathbb{R}^2 \ : \ a_{1,1}x_1 + a_{1,2}x_2 = b_1\}$

$\mathcal{S}_2 = \{\mathbf{x} \in \mathbb{R}^2 \ : \ a_{2,1}x_1 + a_{2,2}x_2 = b_2\}$

$$\Rightarrow \quad \mathbf{x} = \mathcal{S}_1 \cap \mathcal{S}_2$$

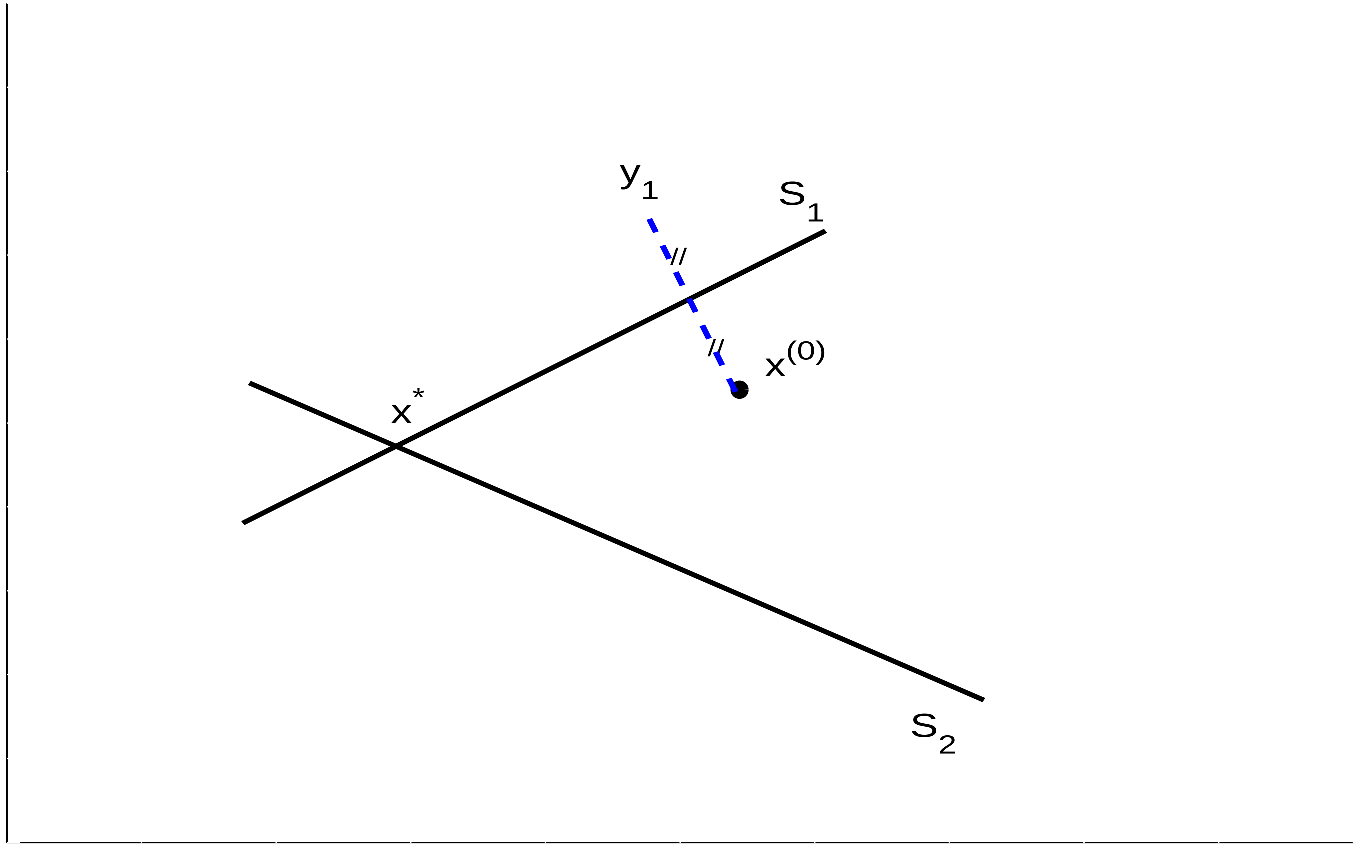
Geometric derivation. II

Initial guess $\mathbf{x}^{(0)}$



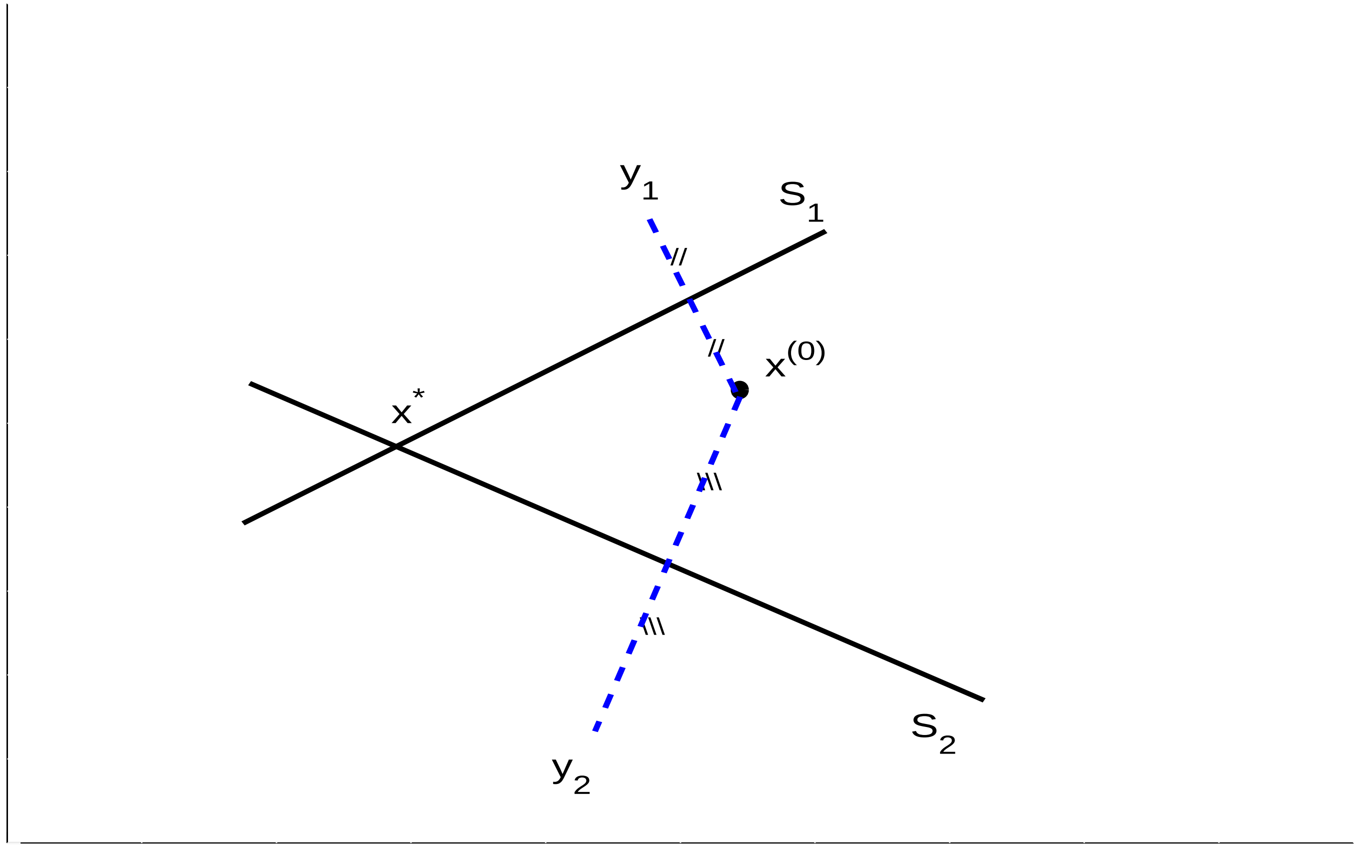
Geometric derivation. II

Initial guess $\mathbf{x}^{(0)}$



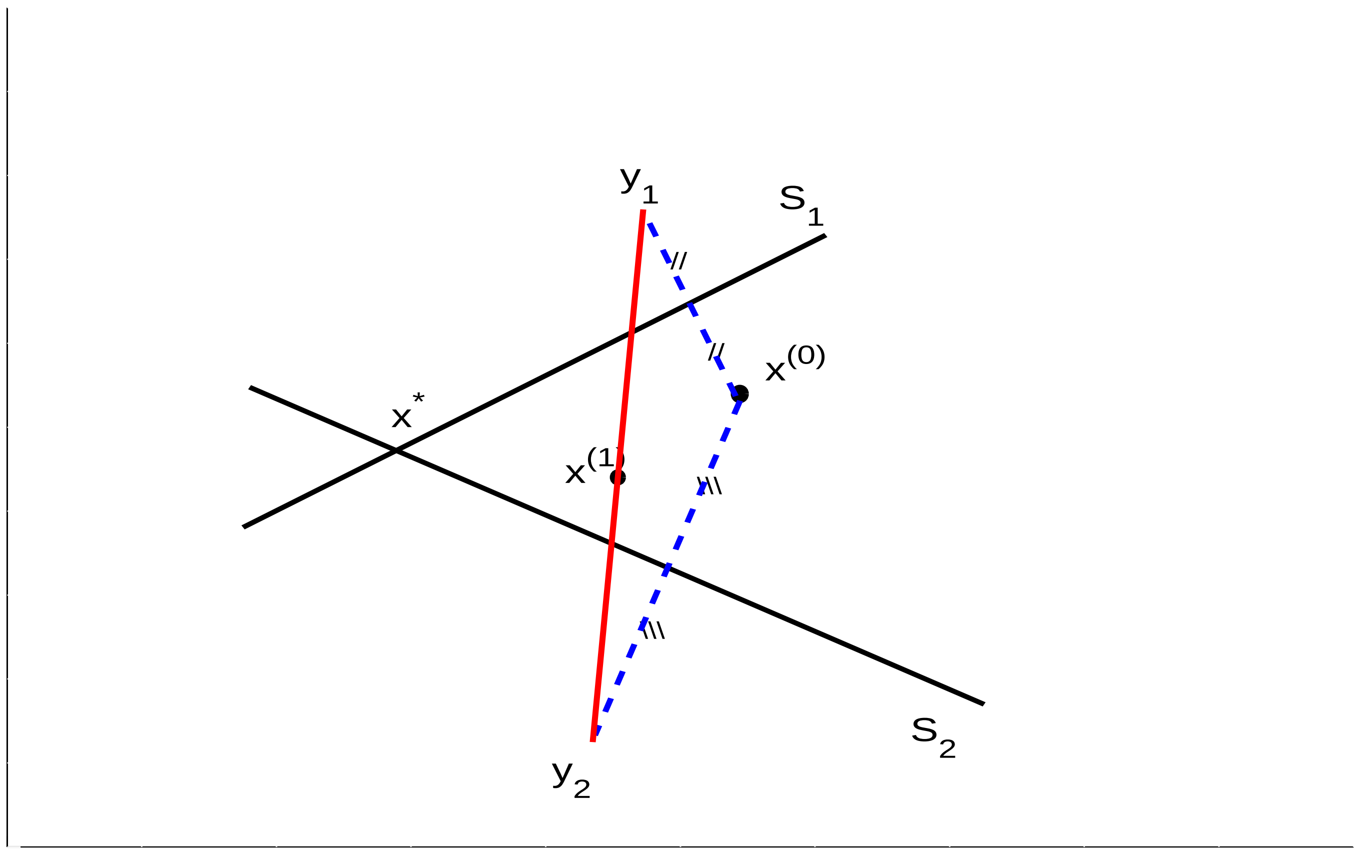
Geometric derivation. II

Initial guess $\mathbf{x}^{(0)}$

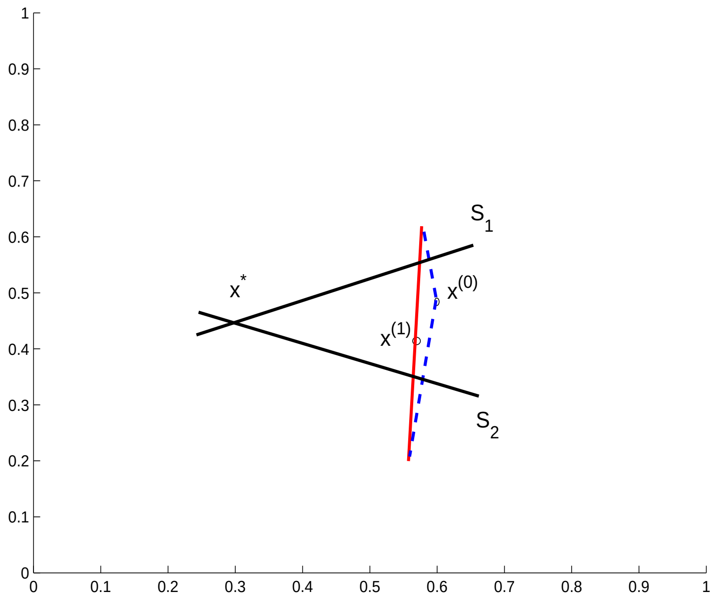


Geometric derivation. II

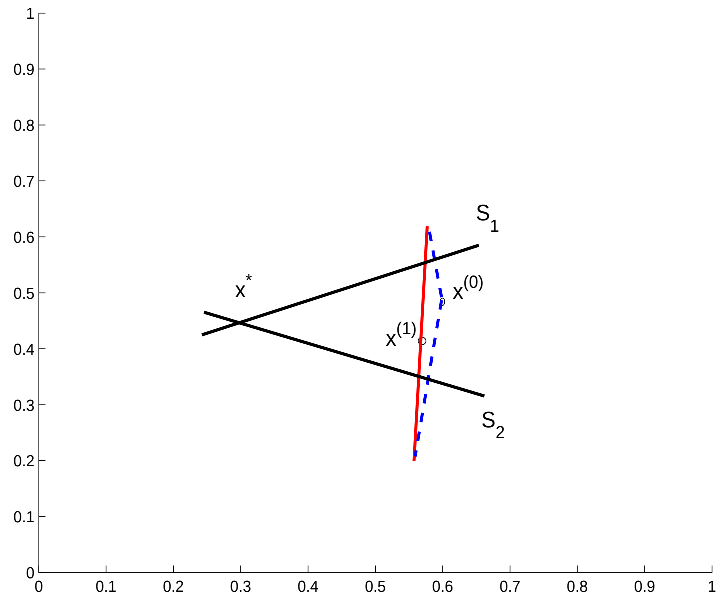
Initial guess $\mathbf{x}^{(0)}$

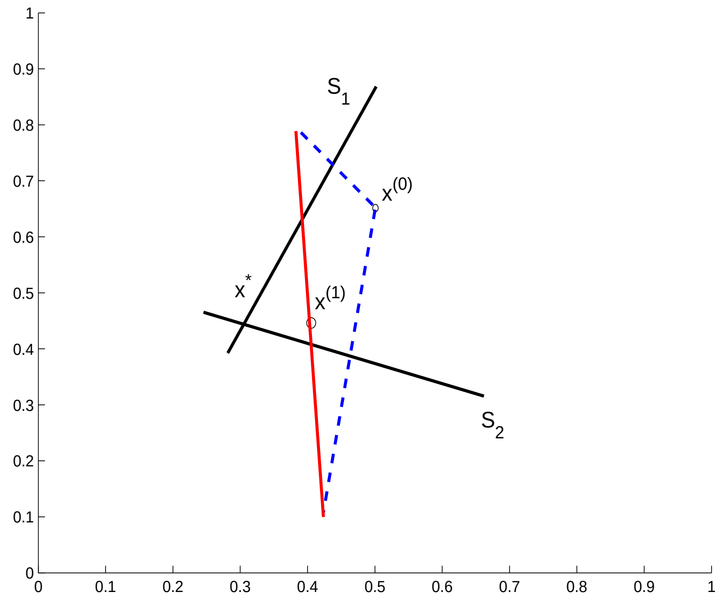


Convergence

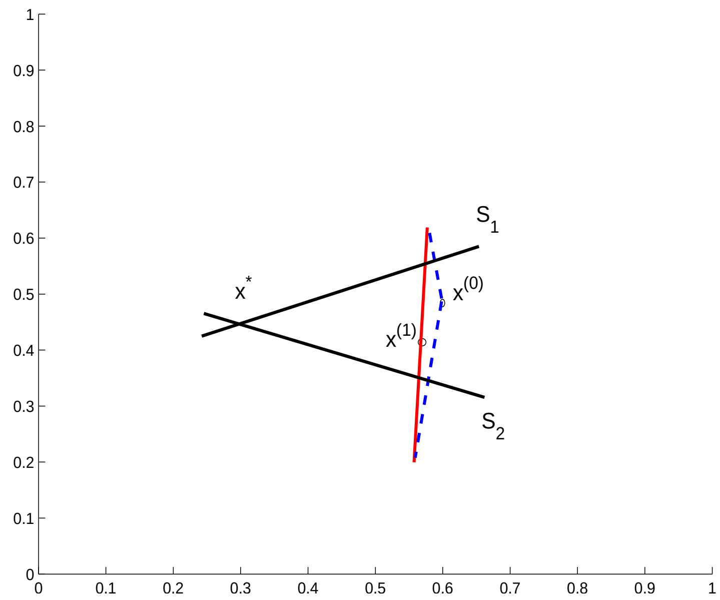


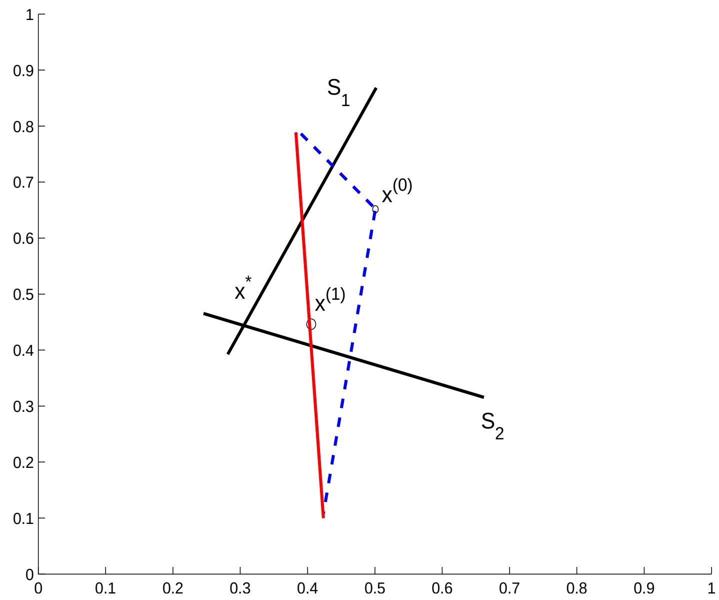
Convergence





Convergence





Linear Convergence. But

The more orthogonal the rows of $\mathbf{A}$, the faster

Note: Convergence depends on spectral radius of sum of scaled proj's.

Family of Methods

- Kaczmarz method (1937)
- Row Projection Methods (see, e.g., R.Bramley)
- ART (Algebraic reconstruction techniques)
- POCS (Projection onto convex sets)

T. Nikazad, Ph.D. Thesis (2008) - cf. T. Elfving.

Algebraic derivation. I

For simplicity of exposition (no loss of generality):

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \dots \end{pmatrix} \text{ rows have unit length}$$

$\Rightarrow \mathbf{a}_i \mathbf{a}_i^T = \mathcal{P}_i$ Orthogonal Projector

Algebraic derivation. I

For simplicity of exposition (no loss of generality):

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \dots \end{pmatrix} \text{ rows have unit length}$$

$\Rightarrow \mathbf{a}_i\mathbf{a}_i^T = \mathcal{P}_i$ Orthogonal Projector

Given initial guess $\mathbf{x}^{(0)}$, $\mathbf{r}^{(0)} = A\mathbf{x}^{(0)} - \mathbf{b}$,

$$\mathbf{y}_i = \mathbf{x}^{(0)} - 2\mathcal{P}_i\mathbf{x}^{(0)}$$

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \omega_1\mathbf{y}_1 + \omega_2\mathbf{y}_2 \\ &= \mathbf{x}^{(k)} - 2\mathcal{P}_1\omega_1(\mathbf{x}^{(k)} - \mathbf{x}^*) - 2\mathcal{P}_2\omega_2(\mathbf{x}^{(k)} - \mathbf{x}^*) \end{aligned}$$

with $\omega_1 + \omega_2 = 1$.

Algebraic derivation. II

Assume $\omega_1 = \omega_2 \equiv \omega$:

$$
\begin{aligned}
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - 2\mathcal{P}_1\omega_1(\mathbf{x}^{(k)} - \mathbf{x}^*) - 2\mathcal{P}_2\omega_2(\mathbf{x}^{(k)} - \mathbf{x}^*) \\
&= \mathbf{x}^{(k)} - 2\omega\mathbf{a}_1\mathbf{a}_1^T(\mathbf{x}^{(k)} - \mathbf{x}^*) - 2\omega\mathbf{a}_2\mathbf{a}_2^T(\mathbf{x}^{(k)} - \mathbf{x}^*) \\
&= \mathbf{x}^{(k)} - 2\omega(\mathbf{a}_1, \mathbf{a}_2)\begin{pmatrix} \mathbf{a}_1^T(\mathbf{x}^{(k)} - \mathbf{x}^*) \\ \mathbf{a}_2^T(\mathbf{x}^{(k)} - \mathbf{x}^*) \end{pmatrix} \\
&= \mathbf{x}^{(k)} - 2\omega\mathbf{A}^T\underbrace{\mathbf{A}(\mathbf{x}^{(k)} - \mathbf{x}^*)}_{\mathbf{A}\mathbf{x}^{(k)}-\mathbf{b}} = \mathbf{x}^{(k)} - 2\omega\mathbf{A}^T\mathbf{r}^{(k)}
\end{aligned}
$$

A Projection method

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - 2\omega \mathbf{A}^T \mathbf{r}^{(k)}, \quad k = 0, 1, 2, \ldots \\ \mathbf{r}^{(k+1)} &= \mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{b} \end{aligned}$$

$\Rightarrow \quad \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \in \mathrm{range}(\mathbf{A}^T)$

- $\mathrm{range}(\mathbf{A}^T)$ contains the exact solution $\mathbf{x}^*$
- But: No global constraint imposed $\Rightarrow$ iterative process

Features and Accelerations

- Linear convergence (with no further hypotheses)

Features and Accelerations

- Linear convergence (with no further hypotheses)
- Block version: Projection onto *groups* of rows

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \\ \vdots \\ \mathbf{A}_j^T \end{pmatrix}$$

(parallelism, data locality)

$\Rightarrow$ Reordering strategies particularly good for $\mathbf{A}$ of small bandwidth

Features and Accelerations

- Linear convergence (with no further hypotheses)
- Block version: Projection onto *groups* of rows

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \\ \vdots \\ \mathbf{A}_j^T \end{pmatrix} \qquad \text{(parallelism, data locality)}$$

$\Rightarrow$ Reordering strategies particularly good for $\mathbf{A}$ of small bandwidth

- Acceleration Procedures acting on $\lambda_k, \Omega = diag(\omega_1, \ldots, \omega_n)$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - 2\lambda_k \mathbf{A}^T \Omega^{-1} \mathbf{r}^{(k)}, \quad k = 0, 1, 2, \ldots$$

$0 < \epsilon_1 \leq \lambda_k \leq 2 - \epsilon_2$

Features and Accelerations

- Linear convergence (with no further hypotheses)
- Block version: Projection onto *groups* of rows

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \\ \vdots \\ \mathbf{A}_j^T \end{pmatrix} \qquad \text{(parallelism, data locality)}$$

$\Rightarrow$ Reordering strategies particularly good for $\mathbf{A}$ of small bandwidth

- Acceleration Procedures acting on $\lambda_k, \Omega = diag(\omega_1, \ldots, \omega_n)$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - 2\lambda_k \mathbf{A}^T \Omega^{-1} \mathbf{r}^{(k)}, \quad k = 0, 1, 2, \ldots$$

$$0 < \epsilon_1 \leq \lambda_k \leq 2 - \epsilon_2$$

- Acceleration Procedures: Conjugate Gradient iteration within the block method

Important generalizations

- ⋆ Rectangular case: $\mathbf{Ax} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{n \times m}, n < m$
- ⋆ Nonlinear equations: $F(\mathbf{x}) = \mathbf{0}$
- ⋆ Inequalities: $\mathbf{Ax} \leq \mathbf{b}, \quad \mathbf{x} \geq 0, \quad \mathbf{A} \in \mathbb{R}^{n \times m}, n < m$
- ⋆ Singular (semidefinite) system: convergence to a weighted least-squares solution that minimizes the weighted sum of the squares distances to the hyperplanes)
- ⋆ Ill-posed Problems

A popular application field

e.g., Censor etal (1980's and later). Jiang & Wang (2001 and later)

Application. radiation therapy treatment planning

Math. Problem. Inverse radiation scattering / Image reconstruction:

$$\text{Find } \mathbf{x} \text{ s.t.} \quad \widehat{\mathbf{b}} \leq \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq 0$$

where

n no. 2D grid points; m no. basis radiation intensity grid points

$\mathbf{A} = (a_{i,j}) \in \mathbb{R}^{n\times m}$, dose of radiation at the jth grid point for the ith intensity distribution grid point

$\mathbf{b}, \widehat{\mathbf{b}}$ permitted and required doses in the patient's cross section

$\mathbf{x}$ acceptable radiation intensity (*the feasible solution*

$\rightarrow$ *Convex Feasibility Problem*)

A popular application field

e.g., Censor etal (1980's and later). Jiang & Wang (2001 and later)

Application. radiation therapy treatment planning

Math. Problem. Inverse radiation scattering / Image reconstruction:

$$\text{Find } \mathbf{x} \text{ s.t. } \quad \widehat{\mathbf{b}} \leq \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq 0$$

Features:

- $\mathbf{A} \in \mathbb{R}^{n \times m}, \; n \gg m$
- Not all rows of $\mathbf{A}$ available at the same time
- $\mathbf{A}$ with small bandwidth

  (only neighboring rays intersect the same pixels)

Perspectives

- Combination of Optimal Projection methods and Geometric approaches

  Some examples in literature. Connection to normal equation

  $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$

- Acceleration techniques for inequalities

- Strategies for cases of rows of $\mathbf{A}$'s *upgrading*