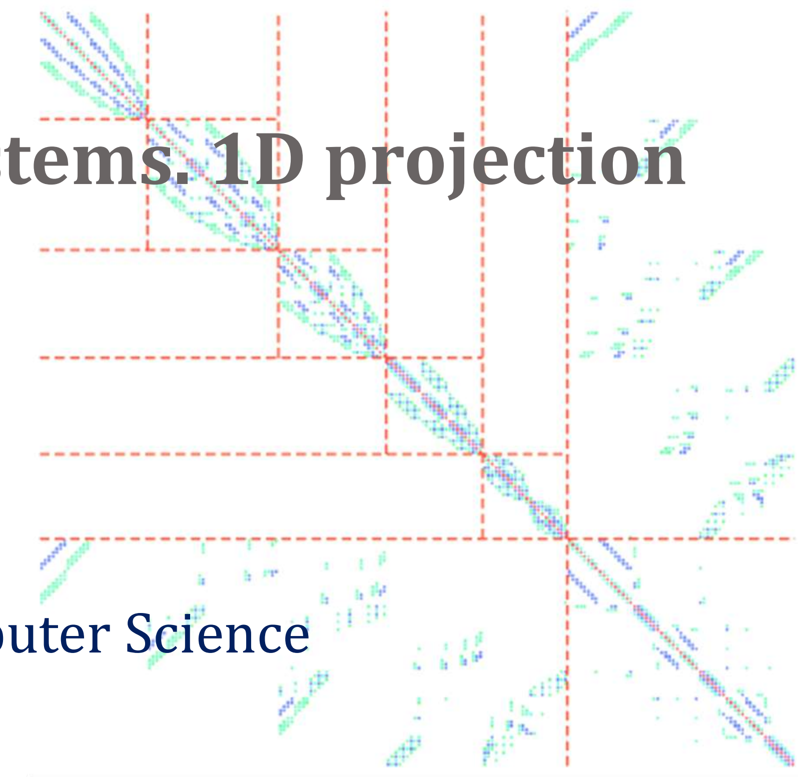
Numerical Methods of Linear Algebra for Sparse Matrices

Lecture 12

Projection methods for linear systems. 1D projection methods



Anna Nasedkina

Department of Mathematical Modeling

Institute of Mathematics, Mechanics and Computer Science

Southern Federal University

Projection methods for linear systems

Idea of projection

General formulation of projection method

1D projection methods

Idea of projection method

- Search the solution $\tilde{x}$ of the system $Ax = b, A \in \mathbb{C}^{n \times n}$, $b \in \mathbb{C}^n$ in the search subspace $K \subset \mathbb{C}^n$
- If $\dim(K) = m$ (the basis of K is defined by m vectors), then it is necessary to impose m constraints to extract approximate solution
- These constraints are defined by the subspace of constraints L, which should also have m vectors in the basis: $\dim(L) = m$
- Typical way to impose constraints is to take orthogonality condition, making residual

  $r = b - A\tilde{x}$ orthogonal to the subspace of constraints L: $r \perp L$ (Petrov-Galerkin condition)

General formulation of a projection method

x_0 is initial guess, K is the search subspace, L is the subspace of constraints

Express approximate solution $\tilde{x}$ as $\tilde{x} = x_0 + \delta$, where δ is correction vector $\delta \in K$

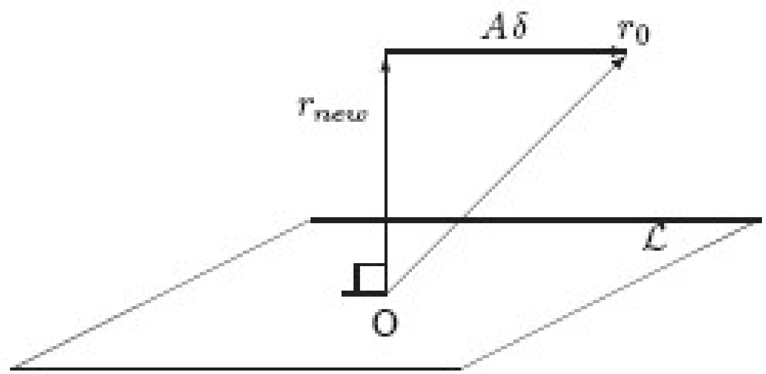
We need to find solution $\tilde{x} \in x_0 + K$, such that residual $r = b - A\tilde{x} \perp L$

Calculate initial residual $r_0 = b - Ax_0$, then

$r = b - A\tilde{x} = b - A(x_0 + \delta) = b - Ax_0 - A\delta = r_0 - A\delta$

$r = r_0 - A\delta \perp L$

$(r_0 - A\delta, w) = 0, \ \forall w \in L$



Matrix representation of a projection method

Consider a linear system $Ax = b,\ A \in \mathbb{C}^{n\times n}$

Size n of the system is large

Take a search subspace $K \subset \mathbb{C}^n,\ \dim(K) = m << n$

$\{v_1, v_2, ..., v_m\}$ is the basis in K

Form a matrix with basis vectors as columns:

$$V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_m \\ | & | & & | \end{pmatrix},\ V \in \mathbb{C}^{n\times m},\ v_i \in \mathbb{C}^n,\ i = \overline{1,m}$$

Take a subspace of constraints $L \subset \mathbb{C}^n,\ \dim(L) = m << n$

$\{w_1, w_2, ..., w_m\}$ is the basis in L

Form a matrix with basis vectors as columns:

$$W = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \dots & w_m \\ | & | & & | \end{pmatrix},\ W \in \mathbb{C}^{n\times m},\ w_i \in \mathbb{C}^n,\ i = \overline{1,m}$$

Matrix representation of a projection method

Approximate solution $\tilde{x}$ is given by $\tilde{x} = x_0 + \delta,\ \delta \in K \subset \mathbb{C}^n$

Let's represent correction vector δ as a linear combination of basis vectors of K: $\delta = y_1 v_1 + y_2 v_2 + \ldots + y_m v_m = Vy$, where

$y = (y_1, y_2, \ldots, y_m)^T$ is the vector of the coefficients of linear combination

V is the matrix of the basis vectors: $V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_m \\ | & | & & | \end{pmatrix},\ V \in \mathbb{C}^{n \times m},$

$v_i \in \mathbb{C}^n,\ y_i \in \mathbb{C},\ i = \overline{1,m} \ \Rightarrow \ \tilde{x} = x_0 + \delta = x_0 + Vy$

Orthogonality condition $(r_0 - A\delta, w) = 0\ \forall w \in L$ is equivalent to the condition for the basis vectors of L:

$(r_0 - A\delta, w_i) = 0;\ \{w_1, w_2, \ldots, w_m\}$ is the basis in L

$(r_0 - AVy, w_i) = 0$

Matrix representation of a projection method

$(r_0 - AVy, w_i) = 0,\ w_i \in \mathbb{C}^n,\ i = \overline{1, m}$

Recall that an inner product can be written as matrix product: $(a, b) = b^H a$

Matrix of basis vectors in L: $W = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \dots & w_m \\ | & | & & | \end{pmatrix},\ W \in \mathbb{C}^{n\times m}$

Hence inner product $(r_0 - AVy, w_i) = 0$ in matrix form is:

$W^H (r_0 - AVy) = 0$

$W^H r_0 - W^H AVy = 0$

$W^H AVy = W^H r_0$

1) We can find y by solving the system $(W^H AV)y = W^H r_0$

Formally $y = (W^H AV)^{-1} W^H r_0$, matrix $W^H AV$ must be nonsingular

2) After finding y we can find the solution x of initial system $Ax = b$ as

$x = x_0 + Vy = x_0 + V(W^H AV)^{-1} W^H r_0$

General projection method: matrix representation and algorithm

- x_0 is initial guess
- $\tilde{x}$ is approximate solution

$\tilde{x} = x_0 + \delta,\ \delta = Vy$, then

$\tilde{x} = x_0 + Vy$

- V is the matrix of the basis vectors in K
- y is the vector of the coefficients to be found
- W is the matrix of the basis vectors in L

$W^H AVy = W^H r_0$

$\tilde{x} = x_0 + V(W^H AV)^{-1} W^H r_0$

Algorithm of general projection method

Take initial guess x_0, calculate initial residual

$r_0 = b - Ax_0$

Until convergence do:

1) Select subspaces K and L, take bases

$$V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_m \\ | & | & & | \end{pmatrix} \text{ in } K$$

$$W = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \dots & w_m \\ | & | & & | \end{pmatrix} \text{ in } L$$

2) Calculate residual $r = b - Ax$, estimate residual norm: $\|r\|_2 < \varepsilon$

3) Find y from solving the system $W^H AVy = W^H r$:

$y = (W^H AV)^{-1} W^H r$

3) Calculate x as $x = x + Vy$

End do

Theorems on general projection method

Matrix $W^H AV$ can be singular, even when A is nonsingular.

Example. Take a block matrix consisting of zero and identity matrices:

$A = \begin{pmatrix} 0 & I \\ I & I \end{pmatrix}$, $0 \in \mathbb{R}^{m \times m}$ is zero matrix, is $I \in \mathbb{R}^{m \times m}$ identity matrix

A is nonsingular. However, if we take the basis with unit orts: $V = W = \{e_1, ..., e_m\}$,

then $W^H AV$ will be singular

Theorem 1. $W^H AV$ is nonsingular $\Leftrightarrow$ there is no $v \in AK$, such that $v \perp L$

Theorem 2. If matrix A is positive definite and $L = K$ or

matrix A is nonsingular and $L = AK$, then $W^H AV$ is nonsingular for any bases W in L, V in K.

Theorem 3. If matrix A is Hermitian (symmetric) positive definite, then $W^H AV$ is Hermitian (symmetric) positive definite.

One-dimensional projection methods

$\dim(K) = \dim(L) = 1$

Then $K = span\{v\}$, vector v is the basis in K

$L = span\{w\}$, vector w is the basis in L

Orthogonality condition: $r_{new} \perp w$ or $(r_{new}, w) = 0$

$r_{new} = r - A\delta$, where δ is correction vector

We had $x_{new} = x + \delta$, $\delta = Vy \Rightarrow x_{new} = x + Vy$

Now $V = v \Rightarrow \delta = \alpha v$, where α is unknown coefficient

$(r_{new}, w) = 0$

$(r - A\delta, w) = 0$

$(r - A\alpha v, w) = 0$

$(r, w) - \alpha(Av, w) = 0$

Hence $\alpha = \dfrac{(r, w)}{(Av, w)}$

General algorithm of 1D projection method

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$

Until convergence (residual becomes small) do:

1. Select v and w as bases for K and L

2. Take approximate solution x, calculate residual $r = b - Ax$,

compute and estimate residual norm: $\|r\|_2 < \varepsilon$

3. Compute coefficient $\alpha = \dfrac{(r, w)}{(Av, w)}$

4. Recompute approximate solution $x = x + \alpha v$

End Do

Examples of 1D projection methods

- Steepest Descent Method
- Minimal Residual Iteration Method
- Residual Norm Steepest Descent Method

Steepest Descent method (SDM)

Matrix $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite

Take $L = K$, $v = w = r$ as the basis in $K = L$

Algorithm of SDM

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$

Until convergence (residual becomes small) do:

1. Calculate residual at current iteration $r_k = b - Ax_k$,

compute and estimate residual norm: $\|r_k\|_2 < \varepsilon$

2. Compute coefficient $\alpha = \dfrac{(r,w)}{(Av,w)} \Rightarrow \alpha_k = \dfrac{(r_k, r_k)}{(Ar_k, r_k)}$

3. Compute new approximate solution $x_{k+1} = x_k + \alpha_k r_k$

End Do

Here $(Ar_k, r_k) > 0$, because A is symmetric positive definite

Two different matrix-by-vector products are computed in this loop:

Ax_k and Ar_k

Modification of SDM algorithm

Algorithm of SDM with one matrix-by-vector product in a loop:

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$, calculate $p_0 = Ar_0$

Until convergence (residual becomes small) do:

1. Compute coefficient $\alpha_k = \dfrac{(r_k, r_k)}{(p_k, r_k)}$

2. Compute new approximate solution $x_{k+1} = x_k + \alpha_k r_k$,

3. Compute new residual $r_{k+1} = r_k - \alpha_k p_k$, estimate residual norm: $\|r_{k+1}\|_2 < \varepsilon$

4. Compute new $p_{k+1} = Ar_{k+1}$

End Do

Here only one matrix-by-vector product is computed in the loop: Ar_{k+1}

Theorems on Steepest Descent method

Theorem 1. The functional $f(x) = \|x - x_*\|_A^2$ is minimized at every iteration in the direction of $-\nabla f(x) \uparrow\uparrow r$ and coefficient $\alpha = \dfrac{(r,r)}{(Ar,r)}$ gives minimum to the functional $f(x) = \|x - x_*\|_A^2$, where is x_* exact solution, r is the residual, $\|x\|_A^2$ is defined as $\|x\|_A^2 = (x,x)_A = (Ax,x)$

Convergence of SDM is guaranted when matrix A is symmetric positive definite.

Theorem 2. If matrix A is symmetric positive definite, then

1) $\left\|x_* - x^{(k+1)}\right\|_A \le \dfrac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \left\|x_* - x^{(k)}\right\|_A$, where $\lambda_{\max}$ and $\lambda_{\min}$ are maximal and minimal eigenvalues of A, x_* is the exact solution of the system $Ax = b$

2) SDM converges for any initial guess

Proof of Theorem 1.1

Theorem 1. The functional $f(x) = \|x - x_*\|_A^2$ is minimized at every iteration in the direction of $-\nabla f(x) \uparrow\uparrow r$ and coefficient $\alpha = \dfrac{(r,r)}{(Ar,r)}$ gives minimum to the functional

$\vartriangle$ Proove that 1) $-\nabla f(x) \uparrow\uparrow r$

Recall that $(x,y)_A = (Ax,y) = y^H Ax$ ($y^T Ax$ in real space $\mathbb{R}^n$, also $(x,y) = (y,x)$ in $\mathbb{R}^n$)

$f(x) = \|x - x_*\|_A^2 = (x - x_*, x - x_*)_A = (A(x - x_*), x - x_*) =$

$= (Ax, x) - (Ax, x_*) - (Ax_*, x_*) - (Ax_*, x) =$

$= (Ax, x) - (Ax, x_*) - (b, x_*) - (b, x) = x^T Ax - x_*^T Ax - b^T x_* - b^T x$

As A is symmetric: $A = A^T$, then $x_*^T Ax = x_*^T A^T x = (Ax_*)^T x = b^T x$

$x^T Ax - x_*^T Ax - b^T x_* - b^T x = x^T Ax - b^T x - b^T x_* - b^T x = x^T Ax - 2b^T x - b^T x_*$

Hence $f(x) = x^T Ax - 2b^T x - b^T x_* = x^T Ax - 2b^T x - const$

$\nabla f(x)$ is the gradient of the functional, we differentiate by x:

$\nabla f(x) = 2Ax - 2b = 2(Ax - b) = -2r$, where $r = b - Ax$ is the residual

Hence $-\nabla f(x) = 2r \;\Rightarrow\; -\nabla f(x) \uparrow\uparrow r$ $\square$

Proof of Theorem 1.2

$\triangle$ Proove that 2) coefficient $\alpha = \dfrac{(r,r)}{(Ar,r)}$ gives minimum to the functional $f(x) = \|x - x_*\|_A^2$:

$$f(x) = \|x - x_*\|_A^2 = x^T Ax - 2b^T x - const$$

Approximate solution x is recomputed in the algrorithm of SDM as $x + \alpha r$

$$f(x + \alpha r) = (x + \alpha r)^T A(x + \alpha r) - 2b^T (x + \alpha r) - const =$$

$$= \underline{\underline{x^T Ax}} + \underline{\alpha x^T Ar} + \alpha^2 r^T Ar + \underline{\alpha r^T Ax} - \underline{\underline{2b^T x}} - 2\alpha b^T r + const =$$

$$= (x^T Ax - b^T x) - b^T x + \underline{\alpha x^T Ar} + \underline{\alpha r^T Ax} - 2\alpha b^T r + \alpha^2 r^T Ar + const$$

$$\underline{\alpha r^T Ax} = \alpha(Ax, r) = \alpha(x, A^T r) = \alpha(x, Ar) = \alpha(A^T x, r) = \alpha(Ax, r) = \underline{\alpha x^T Ar}$$

Hence $f(x + \alpha r) = (x^T Ax - b^T x) - b^T x + 2\alpha x^T Ar - 2\alpha b^T r + \alpha^2 r^T Ar + const$

$$x^T Ax - b^T x = (Ax)^T x - b^T x = ((Ax)^T - b^T)x = (Ax - b)^T x = -r^T x$$

$$2\alpha x^T Ar - 2\alpha b^T r = 2\alpha(x^T Ar - b^T r) = 2\alpha((Ax)^T r - b^T r) = 2\alpha(Ax - b)^T r = -2\alpha r^T r$$

Hence $f(x + \alpha r) = -r^T x - b^T x - 2\alpha r^T r + \alpha^2 r^T Ar + const = \varphi(\alpha)$

$\dfrac{d\varphi(\alpha)}{d\alpha} = -2r^T r + 2\alpha r^T Ar$. In the algorithm of SDM $\alpha = \dfrac{(r,r)}{(Ar,r)} \Rightarrow$

$$\frac{d\varphi(\alpha)}{d\alpha} = -2r^T r + 2\frac{(r,r)}{(Ar,r)} r^T Ar = -2r^T r + 2\frac{r^T r}{r^T Ar} r^T Ar = -2r^T r + 2r^T r = 0$$

Hence $\alpha = \dfrac{(r,r)}{(Ar,r)}$ gives minimum to the functional $\square$

Minimal Residual Iteration method (MRIM)

Matrix $A \in \mathbb{R}^{n\times n}$ is positive definite

Take $L = AK$, $v = r$ is the basis in K, $w = Av = Ar$ is the basis in L

Algorithm of MRIM

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$

Until convergence (residual becomes small) do:

1. Calculate residual at current iteration $r_k = b - Ax_k$,

compute and estimate residual norm: $\|r_k\|_2 < \varepsilon$

2. Compute coefficient $\alpha = \dfrac{(r, w)}{(Av, w)} \Rightarrow \alpha_k = \dfrac{(Ar_k, r_k)}{(Ar_k, Ar_k)}$

3. Compute new approximate solution $x_{k+1} = x_k + \alpha_k r_k$

End Do

Here $(Ar_k, Ar_k) > 0$, because A is positive definite:

$(Ar_k, Ar_k) = (r_k, A^T Ar_k) = (A^T Ar_k, r_k) > 0$, as $A^T A$ is symmetric positive definite.

Two different matrix-by-vector products are computed in this loop:

Ax_k and Ar_k

Modification of MRIM algorithm

Algorithm of MRIM with one matrix-by-vector product in a loop:

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$, calculate $p_0 = Ar_0$

Until convergence (residual becomes small) do:

1. Compute coefficient $\alpha_k = \dfrac{(p_k, r_k)}{(p_k, p_k)}$

2. Compute new approximate solution $x_{k+1} = x_k + \alpha_k r_k$,

3. Compute new residual $r_{k+1} = r_k - \alpha_k p_k$, estimate residual norm: $\|r_{k+1}\|_2 < \varepsilon$

4. Compute new $p_{k+1} = Ar_{k+1}$

End Do

Here only one matrix-by-vector product is computed in the loop: Ar_{k+1}

Theorems on Minimal Residual Iteration method

Theorem 1. The functional $f(x) = \|b - Ax\|_2^2 = \|r\|_2^2$ is minimized at every iteration in the direction of r, where r is the residual.

Convergence of MRIM is guaranted when matrix A is real positive definite.

Theorem 2. If matrix A is real positive definite, then

1) $\|r_{k+1}\|_2 \leq \sqrt{1 - \frac{\mu^2}{\sigma^2}} \|r_k\|_2$, where $\mu = \frac{1}{2}\lambda_{\min}$ $\lambda_{\min}$ is the minimal eigenvalue of $A + A^T$, and $\sigma = \|A\|_2$

2) MRIM converges for any initial guess

Residual Norm Steepest Descent method (RNSD)

Matrix $A \in \mathbb{R}^{n\times n}$ is nonsingular

Take $L = AK$, $v = A^T r$ is the basis in K, $w = Av$ is the basis in L

In general $\alpha = \dfrac{(r,w)}{(Av,w)} \Rightarrow$ here $\alpha_k = \dfrac{(r_k, AA^T r_k)}{(Av_k, Av_k)} = \dfrac{(A^T r_k, A^T r_k)}{(Av_k, Av_k)} = \dfrac{(v_k, v_k)}{(Av_k, Av_k)} = \dfrac{\|v_k\|_2^2}{\|Av_k\|_2^2}$

Algorithm of RNSD

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$

Until convergence (residual becomes small) do:

1. Calculate residual at current iteration $r_k = b - Ax_k$,

compute and estimate residual norm: $\|r_k\|_2 < \varepsilon$

2. Compute $v_k = A^T r_k$

2. Compute coefficient $\alpha = \dfrac{(r,w)}{(Av,w)} \Rightarrow \alpha_k = \dfrac{\|v_k\|_2^2}{\|Av_k\|_2^2}$

3. Compute new approximate solution $x_{k+1} = x_k + \alpha_k r_k$

End Do

Three different matrix-by-vector products are computed in this loop: Ax_k, $A^T r_k$, Av_k

Modification of RNSD algorithm

Algorithm of MRIM with two matrix-by-vector products in a loop:

Take initial guess x_0, calculate initial residual $r_0 = b - Ax_0$

Until convergence (residual becomes small) do:

1. Compute $v_k = A^T r_k$, $p_k = Ar_k$

2. Compute coefficient $\alpha_k = \dfrac{\|v_k\|_2^2}{\|Av_k\|_2^2}$

3. Compute new approximate solution $x_{k+1} = x_k + \alpha_k r_k$,

4. Compute new residual $r_{k+1} = r_k - \alpha_k p_k$, estimate residual norm: $\|r_{k+1}\|_2 < \varepsilon$

End Do

Here only two matrix-by-vector products are computed in the loop: $A^T r_k$ and Ar_k

Theorems on Residual Norm Steepest Descent method

Theorem 1. The functional $f(x) = \|b - Ax\|_2^2 = \|r\|_2^2$ is minimized at every iteration in the direction of $-\nabla f(x)$, where r is the residual.

Convergence of RNSD is guaranted when matrix A is nonsingular.

Theorem 2. If matrix A is nonsingular, then RNSD converges for any initial guess

RNSD method is SDM applied to the system $A^T Ax = A^T b$

Theorems on orthogonal and oblique projection methods

- Orthogonal projection method: $L = K$
- Oblique projection method: $L \neq K$

Theorem 1. If matrix A is symmetric positive definite and $L = K$, then the vector of approximate solution $\tilde{x}$ is the result of orthogonal projection method on K $\Leftrightarrow$ $\tilde{x}$ minimizes the functional $f(x) = \|x - x_*\|_A^2$ on K

Theorem 2. If matrix A is nonsingular and $L = AK$, then, then the vector of approximate solution $\tilde{x}$ is the result of oblique projection method on K $\Leftrightarrow$ $\tilde{x}$ minimizes the functional $f(x) = \|b - Ax\|_2^2 = \|r\|_2^2$ on K