Krylov-Simplex method to solve inverse problems in ℓ_1-norm and max-norm.

Wim Vanroose and Jeffrey Cornelis
U. Antwerpen, Belgium

Sparse Days 2020

Motivation

	original	projected problem	solution
GMRES	$\min \lVert Ax - b \rVert_2$	$\min \lVert H_{k+1,k} y - \lVert r_0 \rVert_2 e_1 \rVert_2$	givens rotations
CG	$\min \lVert e \rVert_A$	$T_{k,k} y = \lVert r_0 \rVert_2 e_1$	recurrences
Krylov	$\lVert Ax - b \rVert_\infty$, $\lVert Ax - b \rVert_1$	?	?

Problem statement

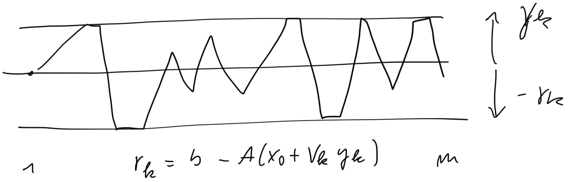
Definition

Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$ a right hand side. The iteraties of the **max-norm Krylov** are given by

$$x_k := \text{argmin}_{x\in x_0+\mathcal{K}_k(A^TA, A^Tr_0)} \max_{i=\{1,\ldots,m\}} |(r_k)_i|. \tag{1}$$

where $r_k = b - A(x_0 + V_k y_k)$ and V_k is a basis for $\mathcal{K}_k(A^TA, A^Tr_0)$

Problem as a LP problem



$$
\begin{aligned}
&\min \gamma_k \\
&\qquad AV_k y_k - r_0 \geq -\gamma_k \\
&\qquad AV_k y_k - r_0 \leq \gamma_k \\
&\qquad \gamma_k \geq 0
\end{aligned}
\tag{2}
$$

Reformulation of LP

$$\begin{aligned}
\min\ & \gamma_k \\
& \gamma_k - r_0 + AV_k y_k \geq 0 \\
& \gamma_k + r_0 - AV_k y_k \geq 0 \\
& \gamma_k \geq 0.
\end{aligned} \tag{3}$$

or

$$\begin{aligned}
\min\ & \gamma_k \\
& -AV_k y_k - \gamma_k \leq -r_0 \\
& AV_k y_k - \gamma_k \leq r_0 \\
& \gamma_k \geq 0.
\end{aligned} \tag{4}$$

or, in matrix notation.

$$\begin{aligned}
\min_{\gamma_k, y_k}\ & \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} \\
& \begin{pmatrix} -1 & -AV_k \\ -1 & AV_k \end{pmatrix} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} \leq \begin{pmatrix} -r_0 \\ r_0 \end{pmatrix}. \\
& \gamma_k \geq 0
\end{aligned} \tag{5}$$

Dual

Lemma

The dual problem of (5) *is*

$$\min \begin{pmatrix} -r_0 & r_0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

$$\begin{pmatrix} -1 & -1 \\ -V^T A^T & V^T A^T \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = - \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{6}$$

$$\lambda \geq 0 \quad \mu \geq 0.$$

There are $k + 1$ conditions and $2N$ unknowns. We know from complementarity condition that only $k + 1$ Lagrange multipliers will differ from zero.

Revised Simplex

An LP in the standard form is

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax = b \\ &\quad\; l \leq x \leq u \end{aligned} \tag{7}$$

where:

- $A \in \mathbb{R}^{m \times n}$ is full rank, $b \in \mathbb{R}^m$
- c, x,l, u are n-vectors.

When A is full rank, there is a collection of m columns that form a non-singular submatrix B.
Indices of selected columns are denoted by:

- $\mathcal{B}$,
- $\mathcal{N}$ remaining indices

Bartels-Golub, Forrest-Tomlin, Reid, ...

Simplex applied to projected system

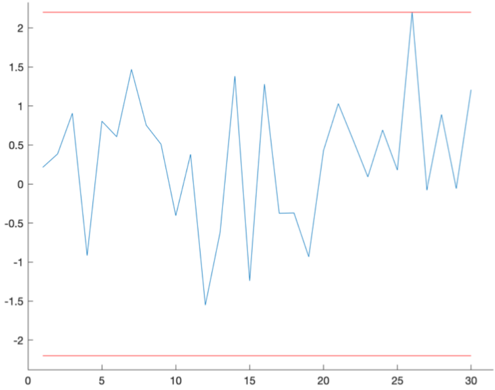
$$
\begin{aligned}
&\min_{\gamma_k, y_k} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} \\
&\begin{pmatrix} -1 & -AV_k \\ -1 & AV_k \end{pmatrix} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} + \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} -r_0 \\ r_0 \end{pmatrix}, \\
&\gamma_k \geq 0, s \geq 0, t \geq 0
\end{aligned}
\tag{8}
$$

Leads to a simplex with $2N$ variables.

Conventions

- Lower bounds in indices: $\{1, \ldots, N\}$
- upper bounds in indices: $\{N+1, \ldots, 2N\}$
- Set of active constraints: $\mathcal{B}_k \subset \{1, \ldots, 2N\}$ for $\mathcal{K}_k$
- Optimal set of active constraints: $\mathcal{B}_k^* \subset \{1, \ldots, 2N\}$ for $\mathcal{K}_k$
- Set of inactive constraints: $\mathcal{N}_k \subset \{1, \ldots, 2N\}$

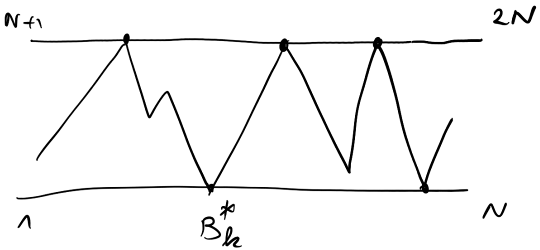
Initial step



where

$$\gamma_0 = \max_i |(r_0)_i| \tag{9}$$

Active Set

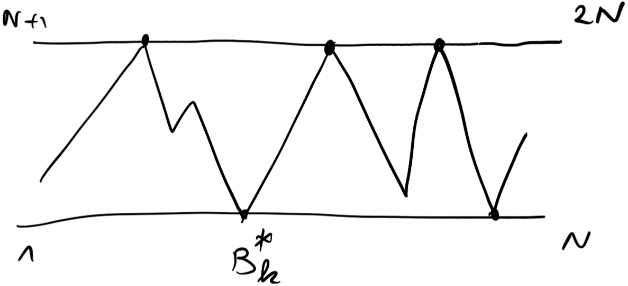


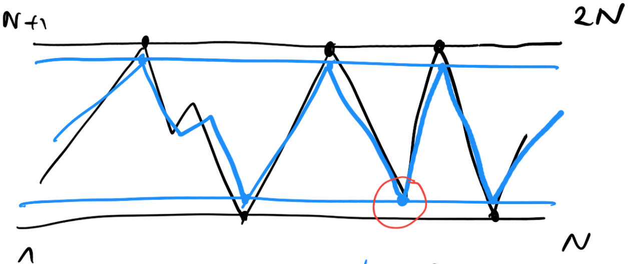
$$\left|\mathcal{B}^*{}_k\right| = k + 1 \tag{10}$$

$$\begin{pmatrix} -1 & -AV_k \\ -1 & AV_k \end{pmatrix}_{i\in\mathcal{B}_k^*} \begin{pmatrix} \gamma_k^* \\ y_k^* \end{pmatrix} = \begin{pmatrix} (-r_0) \\ (r_0) \end{pmatrix}_{i\in\mathcal{B}_k^*} \tag{11}$$

$$\begin{pmatrix} -1 & -AV_k \\ -1 & AV_k \end{pmatrix}_{i\in\mathcal{N}_k^*} \begin{pmatrix} \gamma_k^* \\ y_k^* \end{pmatrix} \leq \begin{pmatrix} (-r_0) \\ (r_0) \end{pmatrix}_{i\in\mathcal{N}_k^*} \tag{12}$$

Expanding the Krylov subspace





$B_{k+1} = B_k^* \cup \{i_r\}$

Initial basic Feasible guess for $\mathcal{K}_{k+1}(A, r_0)$

Definition

We define an **initial basic feasible guess** for iteration $k+1$ with basis V_{k+1}. It is the solution the following auxiliary problem

$$
\begin{aligned}
\min_{\alpha,\gamma_{k+1},y_k^+} \; & \gamma_{k+1}, \\
\text{s.t.} \; & |A(x + V_k y_k^+ + v_{k+1}\alpha) - b|_{i \in \mathcal{B}_k^*} = \gamma_{k+1}, \\
& |A(x + V_k y_k^+ + v_{k+1}\alpha) - b|_{i \in \mathcal{N}_k^*} \leq \gamma_{k+1},
\end{aligned}
\tag{13}
$$

where $\mathcal{B}_k^*$ is the optimal active set for Krylov subspace V_k, the previous step of the algorithm.

One-dimensional subspace

Previous solution γ_k^*, $(y_k^*, 0)$, from V_k, is a feasible solution of the auxiliary problem in V_{k+1}
We change γ_{k+1}, $y_k^+ \in \mathbb{R}^k$ and α, $k+2$ variables, while we satisfy the $k+1$ equations

$$\begin{pmatrix} -1 & -AV_k & -Av_{k+1} \\ -1 & AV_k & Av_{k+1} \end{pmatrix}_{i \in \mathcal{B}_k} \begin{pmatrix} \gamma_{k+1} \\ y_k^+ \\ \alpha \end{pmatrix} = \begin{pmatrix} (-r_0) \\ (r_0) \end{pmatrix}_{i \in \mathcal{B}_k^*} \tag{14}$$

This now defines a one-dimensional subspace that we can parametrise through γ.

Search Direction

We define the matrix

$$B_{k+1} := \begin{pmatrix} -AV_k & -Av_{k+1} \\ AV_k & Av_{k+1} \end{pmatrix} \in \mathbb{R}^{k+1 \times k+1} \tag{15}$$

that allows us to rewrite the system (14) as

$$\begin{aligned} B_{k+1} \begin{pmatrix} y_k^+ \\ \alpha \end{pmatrix} &= \begin{pmatrix} -r_0 \\ r_0 \end{pmatrix}_{i \in \mathcal{B}_k^*} + \gamma_k^* + \Delta\gamma_{k+1} \\ &= \underbrace{\begin{pmatrix} -r_0 \\ r_0 \end{pmatrix}_{i \in \mathcal{B}_k^*} + \gamma_k^*}_{= B_{k+1} \begin{pmatrix} y_k^* \\ 0 \end{pmatrix}} + \Delta\gamma_{k+1} \end{aligned} \tag{16}$$

from which find that

$$\begin{pmatrix} y_k^+ \\ \alpha \end{pmatrix} = \begin{pmatrix} y_k^* \\ 0 \end{pmatrix} + B_{k+1}^{-1} \Delta\gamma_{k+1} = \begin{pmatrix} y_k^* \\ 0 \end{pmatrix} + d\Delta\gamma_{k+1} \tag{17}$$

where $d \in \mathbb{R}^{k+1}$ is the search direction and $\Delta\gamma_{k+1}$ is the step size.

How large is the step size in the search direction?

We start with the lowerbound constraints from the optimization problem. For each index i, we can calculate the step size $\Delta\gamma$

$$(-\gamma_k - \Delta\gamma_{k+1} - AV_k y_k^+)_i = (-r_0)_i \tag{18}$$

reorganisation leads to

$$(\Delta\gamma_{k+1}^{(1)})_i = \frac{(r_0 - \gamma_k - AV_k y_k)_i}{(1 + AV_k d)_i} \quad \text{for} \quad i = 1, \dots, N. \tag{19}$$

Similarly for the upperbound constraints

$$(-\gamma_k - \Delta\gamma_k + AV_k y_k + AV_k d \Delta\gamma_k)_i = (r_0)_i \tag{20}$$

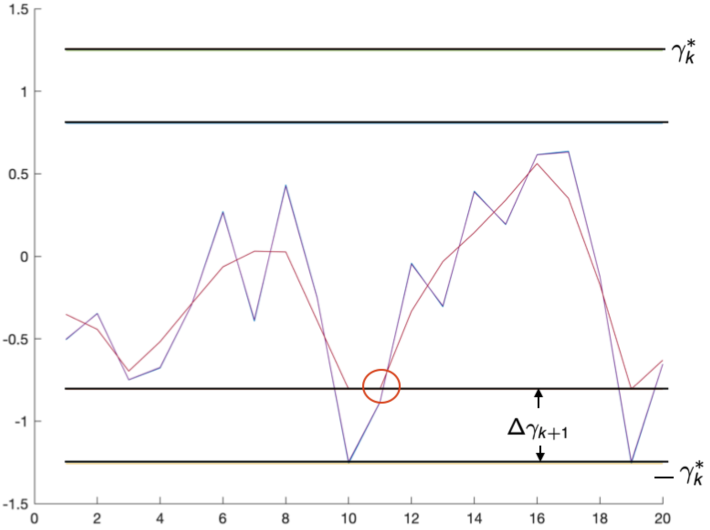
where we find

$$(\Delta\gamma^{(2)})_i = \frac{(-r_0 - \gamma_k + AV_k y_k)_i}{(1 - AV_k d)_i} \quad \text{for} \quad i = 1, \dots, N. \tag{21}$$

The smallest negative value of $\Delta\gamma$ is then

$$\boxed{\Delta\gamma_{k+1} := \max\left(\max_{\Delta\gamma_i < 0, i \notin \mathcal{B}_q} \Delta\gamma_i^1, \max_{\Delta\gamma_i < 0, i \notin \mathcal{B}_q} \Delta\gamma_i^2\right)} \tag{22}$$

Initial basic feasible guess



$$\mathcal{B}_{k+1} = \mathcal{B}_k^* \cup \{r\} \tag{23}$$

Optimal basic set $\mathcal{B}_{k+1}$?

Dual conditions from the KKT for the subset of non-zero Lagrange multipliers

$$\begin{aligned} -\sum \mu_k - \lambda_k &= -1, \\ V^T A^T \lambda_k - V^T A^T \mu_k &= 0, \\ \lambda \geq 0 \quad \mu &\geq 0. \end{aligned} \tag{24}$$

or in matrix form

$$C_{k+1}^T \begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ V^T A^T & -V^T A^T \end{pmatrix} \begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{25}$$

This is $k+2$ square system that we can solve

$$\begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix}_{i \in \mathcal{B}_{k+1}} = C_{k+1}^{-T} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{26}$$

If the solution satisfies $\lambda_k \geq 0$ and $\mu_k \geq 0$, we have the optimal basic set and $\mathcal{B}_{k+1}^* := \mathcal{B}_{k+1}$. Otherwise, pivot as in classical revised Simplex.

Algorithm 1: Krylov-Simplex. Outer $\rightarrow$ Krylov, inner $\rightarrow$ simplex

$r_0 = b - Ax_0$;
$\gamma_0, i = \max_i |(r_0)_i|$;
$\mathcal{B}_0 = \{i\}$ index where the max is reached;
$V_1 = [r_0/\|r_0\|]$;
for $k = 1, \ldots$ **do**
    Calculate $AV_k = [AV_{k-1} Av_k]$ and store. $B_k = \begin{pmatrix} -AV_k \\ AV_k \end{pmatrix}_{i \in \mathcal{B}_{k-1}}$;
    $d_1 = B_k^{-1} 1$;
    $r, \Delta\gamma, y_k$ = blockingfunction $(d_1, \mathcal{B}_k)$;
    $\mathcal{B}_k = \mathcal{B}_{k-1} \cup \{r\}$;
    while ... **do**
        $C_k = \begin{pmatrix} -1 & -AV_k \\ -1 & AV_k \end{pmatrix}_{i \in \mathcal{B}_k}$;
        $\lambda = C_k^{-T} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$;
        if $\lambda_i \geq 0$ **then**
            Break; Solution Found ;
        else
            $q = \min(\lambda_i)$ leaving index;
            $\mathcal{B}_k = \mathcal{B}_k \setminus \{q\}$;
            $d_2 = C_k^{-1} e_q$;
            $r, \Delta\gamma, y_k$ = blockingfunction $(d_2, \mathcal{B}_k)$;
            $\mathcal{B}_k = \mathcal{B}_l \cup \{r\}$;
        end
    end
    $x_k = x_0 + V_k y_k$;
    $\|r_k\|_\infty = \gamma_k$;
    expand $V_{k+1} = [V_k, v_{k+1}]$ using Arnoldi;
end

Max-Norm versus Krylov

Let us recall:

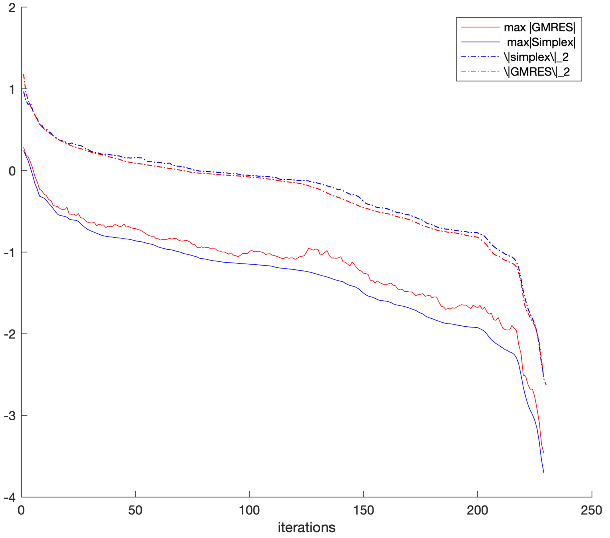
$$\|x\|_\infty = \max_i |x_i| = \max_i \sqrt{|x_i|^2} \leq \sqrt{\sum |x_i|^2} = \|x\|_2 \tag{27}$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \leq \sqrt{n}\|x\|_2 \tag{28}$$

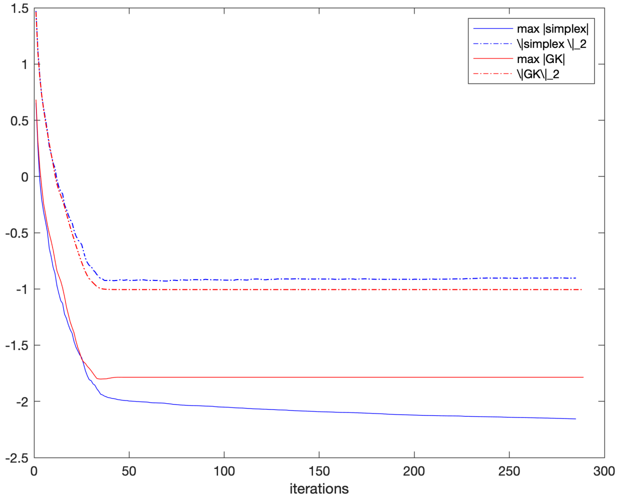
$$\|x\|_2^2 \leq n\|x\|_\infty^2 \tag{29}$$

$$\|x\|_2 \leq \sqrt{n}\|x\|_\infty \tag{30}$$

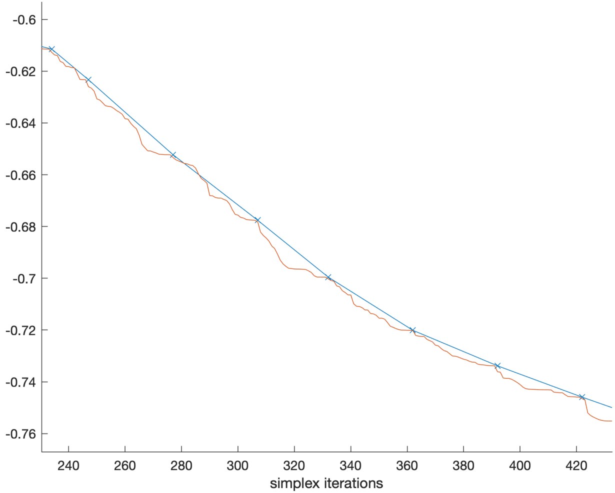
Convergence Krylov-simplex vs GMRES, $A \in \mathbb{R}^{n \times n}$



Convergence Krylov-Simplex vs Golub-Kahan, $A \in \mathbb{R}^{m \times n}$



outer/innerloop



It is beneficial to stop inner loop early and expand the Krylov subspace.

Summary and Conclusions

In progress

- Reuse factorisation: QRupdate, Bartels-Golub for small dense Matrices
- Similar Krylov-Simplex Algorithm for ℓ_1-norm.
- Exploiting the Hessenberg/Tridiagonal structure.
- Analysis of stability and LU factors reuse.

Applications

- Calibration and inverse problems in financial, optical and other complex systems.

Conclusion

- It is possible to solve the projected LP system with simplex.

Happy to collaborate.