Subspace methods for inverse Problems

Generalised Krylov, Krylov-Simplex, Krylov-Newton, Krylov-Active Set, ..

Jeffrey Cornelis, Bas Symoens, Wim Vanroose

Universiteit Antwerpen

Outline

- I: Introduction: What are inverse problems?
- II: Krylov-Simplex for $\min \|Ax - b\|_\infty$
- III: Krylov-Newton for $\min \|Ax - b\|_2 + \lambda \|Lx\|_2$ with automatic choice of λ
- V: Krylov-Active Set to solve $\min \|Ax - b\|_2$ subject to $l \leq x \leq u$.

Universiteit Antwerpen

I: Introduction

Need for regularisation

Consider a linear system of equations

$$Ax = b = \underbrace{b_{\mathrm{ex}}}_{Ax_{\mathrm{ex}}} + \underbrace{e}_{\propto N(0,\sigma)}$$

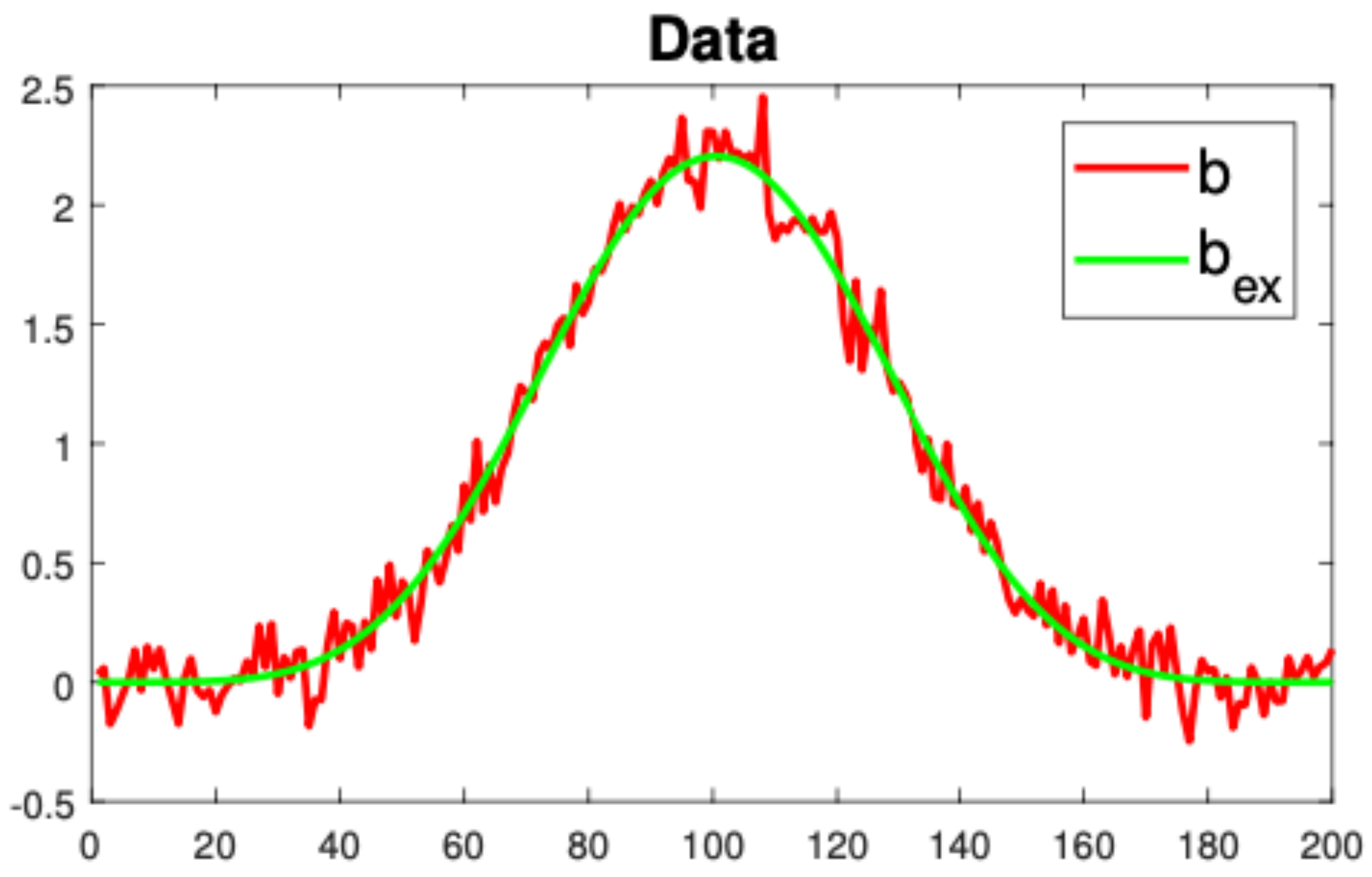
with matrix $A \in \mathbb{R}^{m \times n}$ and data $b \in \mathbb{R}^m$.

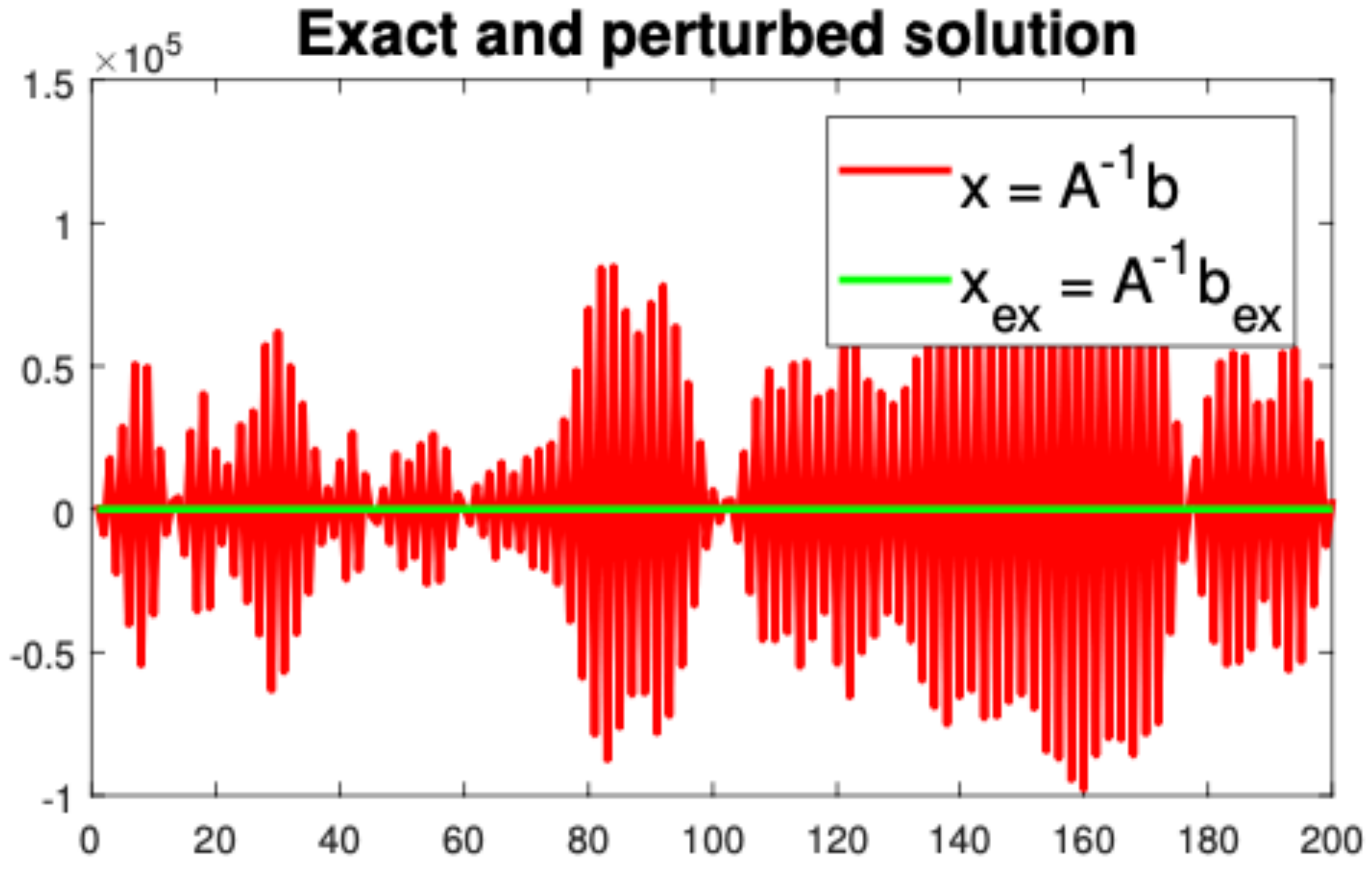
For square nonsingular matrices we have

$$\frac{||\delta x||_2}{||x||_2} \leq \underbrace{||A^{-1}||_2||A||_2}_{=\kappa(A)} \frac{||e||_2}{||b||_2}$$

with $\delta x := x - x_{ex}$.

Small changes in b can lead to huge changes in x for ill-conditioned matrix A.





phillips from RegTools with $\kappa(A) \approx 10^7$.

Universiteit Antwerpen

Tikhonov regularisation

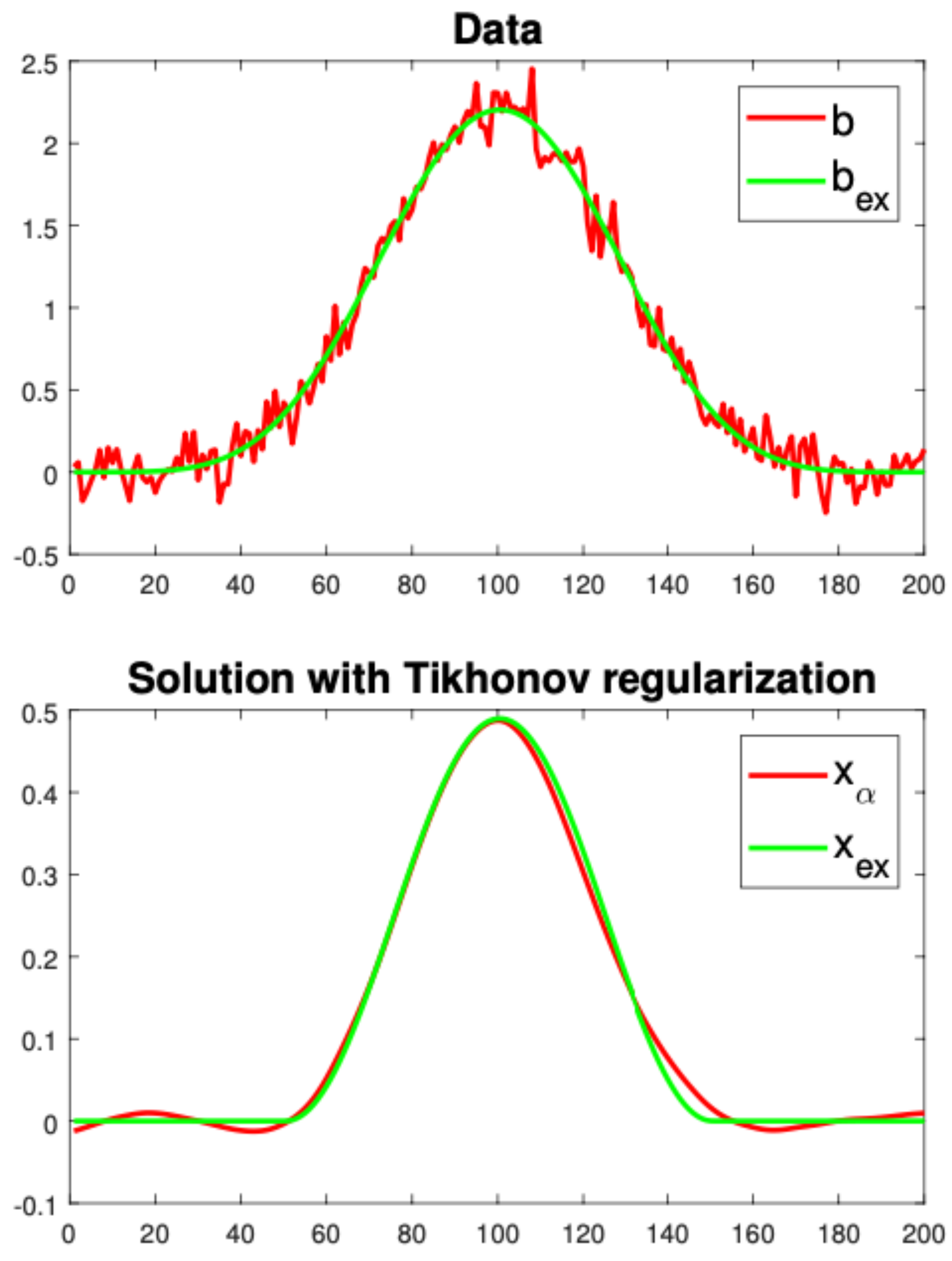
For rectangular matrices ($m \geq n$) we solve the linear least squares problem

$$x = \text{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = (A^T A)^{-1} A^T b$$

Sensitivity now related to $\kappa(A^T A)$.

Tikhonov regularization is a popular approach to obtain a meaningful solution:

$$x = \text{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \alpha \|x\|^2$$
$$= (A^T A + \alpha I)^{-1} A^T b$$



`phillips` from `RegTools` with $\kappa(A) \approx 10^7$.

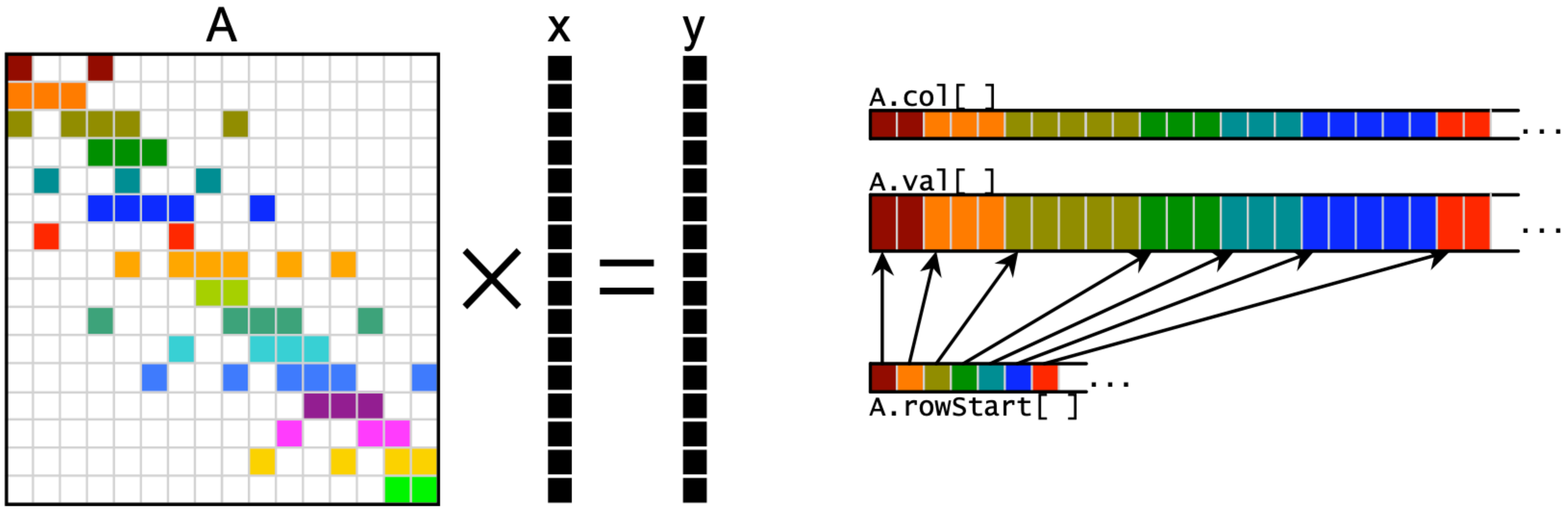
Universiteit Antwerpen

Motivation

- Many inverse problems are formulated as a mixture $|| \cdot ||_2, \ || \cdot ||_\infty, \ || \cdot ||_1$
  - Lasso: $\min ||b - Ax||_2 + \lambda ||Lx||_1$
  - Elastic net: $\min ||b - Ax||_2 + \lambda ||Lx||_1 + \mu ||Mx||_2$
  - Max-norm regularisation $\min ||b - Ax||_2 + \lambda ||Lx||_\infty$
  - Regularisation by bounds $\min ||b - Ax||_2$ subject to $l \leq x \leq u$

What are Krylov Subspace Methods?

$$\mathcal{K}(A, v) = \mathrm{span}\{v, Av, A^2v, \ldots, A^{k-1}v\}$$



Preconditioning = Accelerating Krylov

- Instead of Solving $Ax = b$ solve. $M^{-1}Ax = M^{-1}b$

$$x \in x_0 + \underbrace{span\{v, M^{-1}Av, (M^{-1}A)^2v, \ldots, (M^{-1}A)^{k-1}v\}}_{\textit{Krylov subspace of size k}}$$

- Problem M is any approximation of A
  - Physics based: Simplify the original problem such that it becomes easy to solve
  - Algebraic based: Approximately solve the algebraic equations.

Universiteit Antwerpen

Krylov Subspace methods.

$$\mathscr{K}(A, v) = \text{span}\{v, Av, A^2v, \ldots, A^{k-1}v\}$$

Conjugate Gradients

$$\min_{x \in x_0 + \mathscr{K}(A, r_0)} \|x - x^*\|_A \Rightarrow \quad y_k = T_{k,k}^{-1} \|r_0\|_2 e_1$$

Small tridiagonal linear system.

GMRES

$$\min_{x \in x_0 + \mathscr{K}(A, r_0)} \|b - Ax\|_2 \Rightarrow \quad \min_{y \in \mathbb{R}^k} \left\| \|r_0\|_2 e_1 - H_{k+1,k} y_k \right\|_2$$

Small Least Squares problem

Krylov Subspace methods.

$$\mathcal{K}(A, v) = \text{span}\{v, Av, A^2v, \ldots, A^{k-1}v\}$$

Conjugate Gradients

$$\min_{x \in x_0 + \mathcal{K}(A, r_0)} \|x - x^*\|_A \Rightarrow \quad y_k = T_{k,k}^{-1} \|r_0\|_2 e_1$$

Small tridiagonal linear system.

GMRES

$$\min_{x \in x_0 + \mathcal{K}(A, r_0)} \|b - Ax\|_2 \Rightarrow \quad \min_{y \in \mathbb{R}^k} \left\| \|r_0\|_2 e_1 - H_{k+1,k} y_k \right\|_2$$

Small Least Squares problem

$$\min_{x \in x_0 + \mathcal{K}(A^T A, r_0)} \|b - Ax\|_\infty \Rightarrow$$

??

Krylov Subspace methods.

$$\mathcal{K}(A, v) = \text{span}\{v, Av, A^2v, \ldots, A^{k-1}v\}$$

Conjugate Gradients

$$\min_{x \in x_0 + \mathcal{K}(A, r_0)} \|x - x^*\|_A \Rightarrow \quad y_k = T_{k,k}^{-1} \|r_0\|_2 e_1$$

Small tridiagonal linear system.

GMRES

$$\min_{x \in x_0 + \mathcal{K}(A, r_0)} \|b - Ax\|_2 \Rightarrow \quad \min_{y \in \mathbb{R}^k} \left\| \|r_0\|_2 e_1 - H_{k+1,k} y_k \right\|_2$$

Small Least Squares problem

$$\min_{x \in x_0 + \mathcal{K}(A^TA, r_0)} \|b - Ax\|_\infty \Rightarrow$$

??

$$\min_{x_0 + \mathcal{K}(A^TA, r_0)} \|Lx\|_2^2 \quad \text{s.t. } \|Ax - b\|_2^2 = \|e\|_2^2 \Rightarrow$$

??

Krylov Subspace methods.

$$\mathscr{K}(A, v) = \text{span}\{v, Av, A^2v, \ldots, A^{k-1}v\}$$

Conjugate Gradients

$$\min_{x \in x_0 + \mathscr{K}(A, r_0)} \|x - x^*\|_A \Rightarrow \quad y_k = T_{k,k}^{-1} \|r_0\|_2 e_1$$

Small tridiagonal linear system.

GMRES

$$\min_{x \in x_0 + \mathscr{K}(A, r_0)} \|b - Ax\|_2 \Rightarrow \quad \min_{y \in \mathbb{R}^k} \left\| \|r_0\|_2 e_1 - H_{k+1,k} y_k \right\|_2$$

Small Least Squares problem

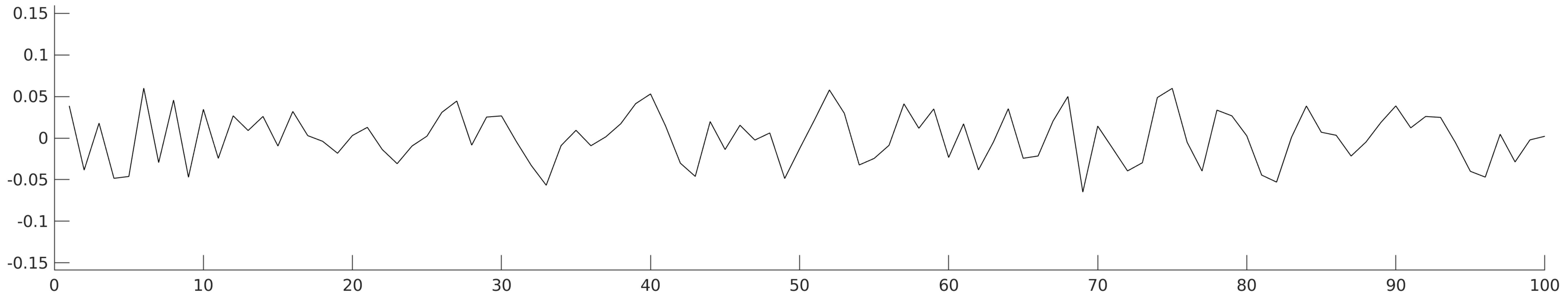
$$\min_{x \in x_0 + \mathscr{K}(A^TA, r_0)} \|b - Ax\|_\infty \Rightarrow \quad ??$$

$$\min_{x_0 + \mathscr{K}(A^TA, r_0)} \|Lx\|_2^2 \quad \text{s.t. } \|Ax - b\|_2^2 = \|e\|_2^2 \Rightarrow \quad ??$$

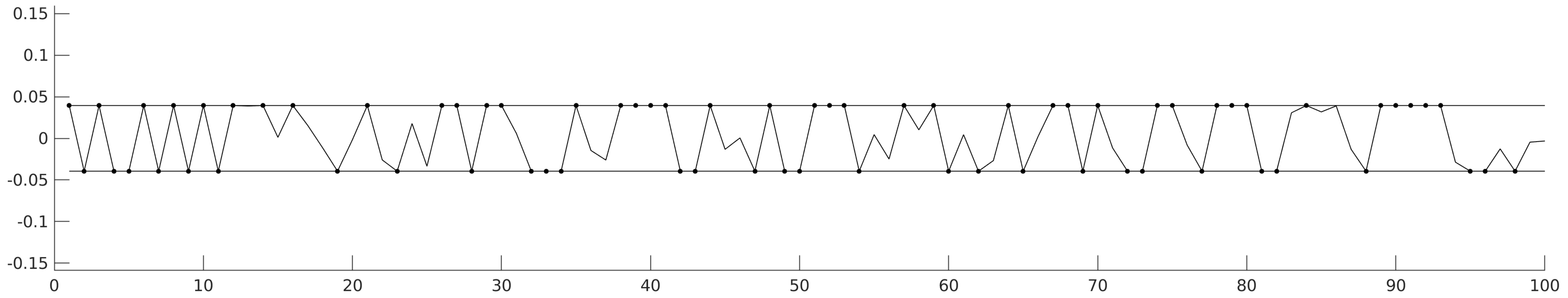
$$\min_{x \in x_0 + \mathscr{K}(A^TA, r_0)} \|Ax - b\|_2^2 \quad \text{s.t. } l \leq x \leq u \Rightarrow \quad ??$$

Universiteit Antwerpen

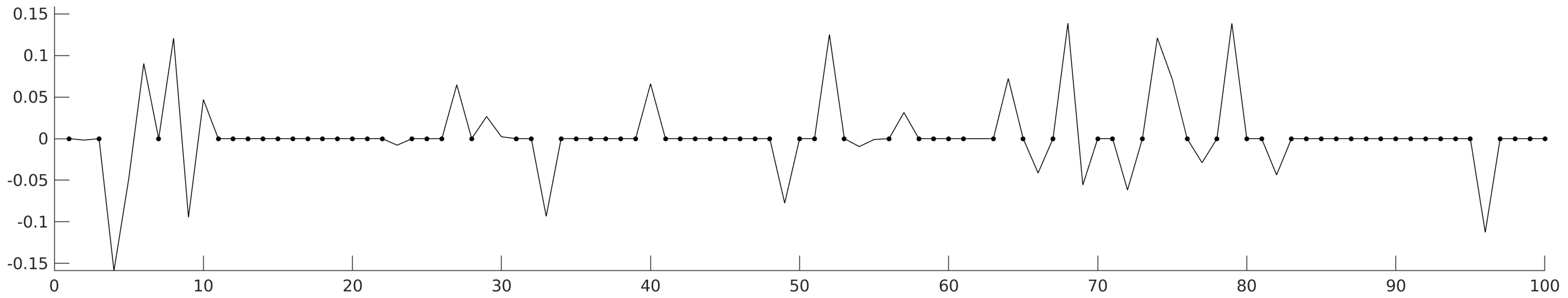
Residual in different norms for $A \in \mathbb{R}^{m \times n}$



$$\min \|b - Ax\|_2$$



$$\min \|b - Ax\|_\infty$$



$$\min \|b - Ax\|_1$$

Universiteit Antwerpen

II: Minimising $\|Ax - b\|_\infty$ over a Krylov subspace: Krylov-Simplex

$$A \in \mathbb{R}^{m \times n}$$

Universiteit Antwerpen

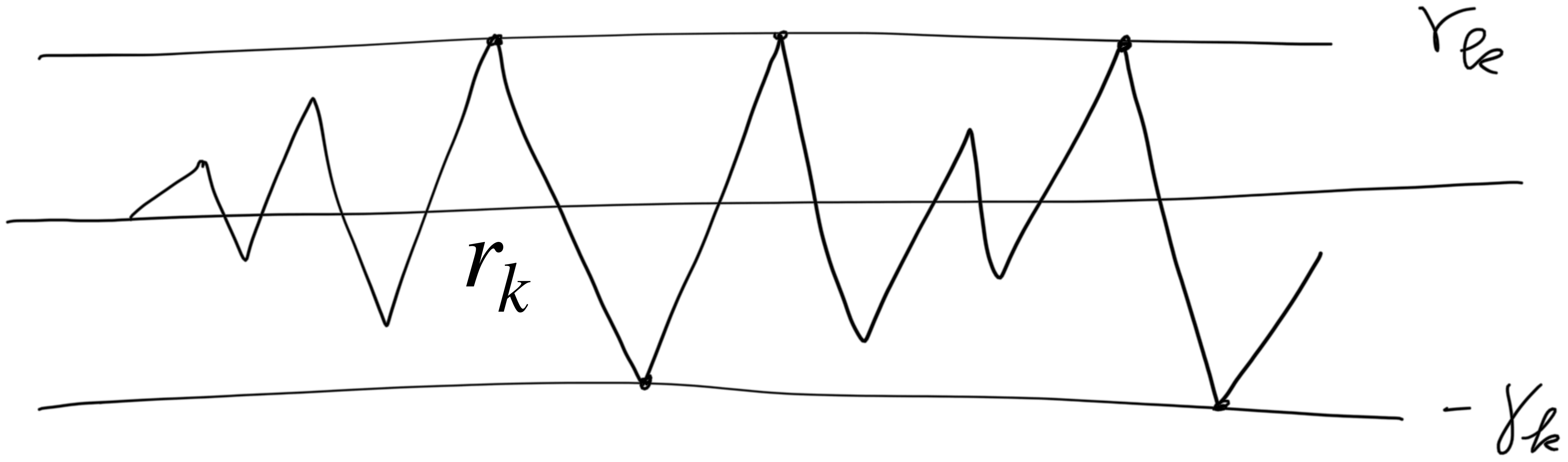
Minimisation of residual in max-norm

Definition $x_k := \text{argmin}_{x \in x_0 + \mathcal{K}(A^TA, r_0)} \|r_k\|_\infty = \min_{y_k \in \mathbb{R}^k} \|r_0 - AV_k y_k\|_\infty$

Formulation as an LP

$$\min_{(\gamma_k, y_k) \in \mathbb{R}^{k+1}} \gamma_k$$

$$\text{s.t.} -\gamma_k \leq (r_0 - AV_k y_k)_j \leq \gamma_k \quad \forall j \in \{1, \ldots, m\}$$



Universiteit Antwerpen

KKT conditions

$$\min_{(\gamma_k, y_k) \in \mathbb{R}^{k+1}} \gamma_k$$

$$\text{s.t.} -\gamma_k \leq (r_0 - AV_k y_k)_j \leq \gamma_k \quad \forall j \in \{1, \ldots, m\}$$

$$\min_{\gamma_k, y} (1 \ 0)^T \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix}$$

$$\begin{pmatrix} -1_m & AV_k \\ -1_m & -AV_k \end{pmatrix} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} \leq \begin{pmatrix} r_0 \\ -r_0 \end{pmatrix} \begin{matrix} \leftarrow \lambda \\ \leftarrow \mu \end{matrix}$$

$$1_m^T(\lambda + \mu) = 1,$$
$$V_k^T A^T(\lambda - \mu) = 0,$$
$$\lambda \geq 0, \quad \mu \geq 0,$$

DUAL FEASIBILITY

$$-\gamma_k 1_m \leq r_0 - AV_k y_k \leq \gamma_k 1_m$$

PRIMAL FEASIBILITY

$$\lambda_j(r_0 - AV_k y + \gamma_k 1_m)_j = 0, \quad \forall j \in \{1, \ldots, m\}$$
$$\mu_j(\gamma_k 1_m - (r_0 - AV_k y))_j = 0, \quad \forall j \in \{1, \ldots, m\}.$$

COMPLEMENTARITY CONDITIONS

Universiteit Antwerpen

Review of revised simplex.

$$\min c^T x$$
$$Ax = b$$
$$l \leq x \leq u$$

$$x_i = \begin{cases} l & i \in \mathscr{L} \\ l < x_i < u & i \in \mathscr{B} \\ u & i \in \mathscr{U} \end{cases}$$

NON-BASIC lower-bound indices

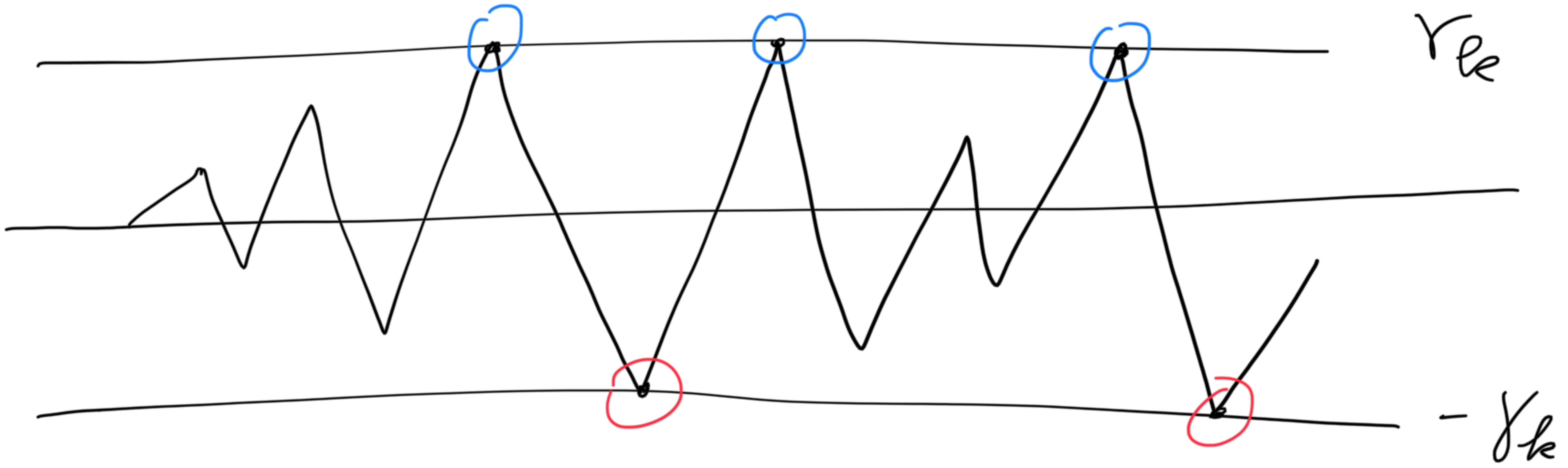
BASIC SET of indices

NON-BASIC upper-bound indices

- The number of element in basic set $|\mathscr{B}| = m$. The non-basis set is $\mathscr{N} = \mathscr{L} \cup \mathscr{U}$.
- $B := A_{I \in \mathscr{B}}$ is the basic matrix is non-singular.

Universiteit Antwerpen

Definition of basic and non-basic set.



$$\mathscr{B} = \mathscr{B}_{<} \cup \mathscr{B}_{>}$$

BASIC SET

$$|\mathscr{B}| = k$$

$$-\gamma_k \leq (r_0 - AV_k y_k)_i \leq \gamma_k \quad \forall i \in \mathscr{N}$$

NON-BASIC SET

Primal and Dual equation

Primal Equation

$$-\gamma_k = (r_0 - AV_k y_k)_i \qquad \forall i \in \mathcal{B}_{k,<}$$

$$(r_0 - AV_k y_k)_i = \gamma_k \qquad \forall i \in \mathcal{B}_{k,>}$$

$$|(r_0 - AV y_k)_i| \leq \gamma_k \qquad \forall i \in \mathcal{N}_k$$

Universiteit Antwerpen

Primal and Dual equation

Primal Equation

$$-\gamma_k = (r_0 - AV_k y_k)_i \qquad \forall i \in \mathscr{B}_{k,<}$$

$$(r_0 - AV_k y_k)_i = \gamma_k \qquad \forall i \in \mathscr{B}_{k,>}$$

$$|(r_0 - AVy_k)_i| \leq \gamma_k \qquad \forall i \in \mathscr{N}_k$$

$$\Rightarrow \underbrace{\begin{pmatrix} -1_{\mathscr{B}_{k,<}} & AV_k|_{\mathscr{B}_{k,<}} \\ 1_{\mathscr{B}_{k,>}} & AV_k|_{\mathscr{B}_{k,>}} \end{pmatrix}}_{:=B_k} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = B_k \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = \begin{pmatrix} r_0|_{\mathscr{B}_{k,<}} \\ r_0|_{\mathscr{B}_{k,>}} \end{pmatrix}$$

Universiteit Antwerpen

Primal and Dual equation

Primal Equation

$$-\gamma_k = (r_0 - AV_k y_k)_i \qquad \forall i \in \mathscr{B}_{k,<}$$
$$(r_0 - AV_k y_k)_i = \gamma_k \qquad \forall i \in \mathscr{B}_{k,>}$$
$$|(r_0 - AVy_k)_i| \leq \gamma_k \qquad \forall i \in \mathscr{N}_k$$

$$\Rightarrow \underbrace{\begin{pmatrix} -1_{\mathscr{B}_{k,<}} & AV_k|_{\mathscr{B}_{k,<}} \\ 1_{\mathscr{B}_{k,>}} & AV_k|_{\mathscr{B}_{k,>}} \end{pmatrix}}_{:=B_k} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = B_k \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = \begin{pmatrix} r_0|_{\mathscr{B}_{k,<}} \\ r_0|_{\mathscr{B}_{k,>}} \end{pmatrix}$$

Dual Equation and complementarity

$$1_m^T(\lambda + \mu) = 1,$$
$$V_k^T A^T(\lambda - \mu) = 0,$$
$$\lambda_j(r_0 - AV_k y + \gamma_k 1_m)_j = 0, \quad \forall j \in \{1,\ldots,m\}$$
$$\mu_j(\gamma_k 1_m - (r_0 - AV_k y))_j = 0, \quad \forall j \in \{1,\ldots,m\}.$$

Primal and Dual equation

Primal Equation

$$-\gamma_k = (r_0 - AV_k y_k)_i \qquad \forall i \in \mathcal{B}_{k,<}$$
$$(r_0 - AV_k y_k)_i = \gamma_k \qquad \forall i \in \mathcal{B}_{k,>}$$
$$|(r_0 - AVy_k)_i| \leq \gamma_k \qquad \forall i \in \mathcal{N}_k$$

$$\Rightarrow \underbrace{\begin{pmatrix} -1_{\mathcal{B}_{k,<}} & AV_k|_{\mathcal{B}_{k,<}} \\ 1_{\mathcal{B}_{k,>}} & AV_k|_{\mathcal{B}_{k,>}} \end{pmatrix}}_{:=B_k} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = B_k \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = \begin{pmatrix} r_0|_{\mathcal{B}_{k,<}} \\ r_0|_{\mathcal{B}_{k,>}} \end{pmatrix}$$

Dual Equation and complementarity

$$1_m^T(\lambda + \mu) = 1,$$
$$V_k^T A^T (\lambda - \mu) = 0,$$

$$\Rightarrow \begin{pmatrix} -1^T_{\mathcal{B}_{k,<}} & 1^T_{\mathcal{B}_{k,>}} \\ V_k^T A^T|_{\mathcal{B}_{k,<}} & V_k^T A|_{\mathcal{B}_{k,>}} \end{pmatrix} \begin{pmatrix} \lambda_{\mathcal{B}_{k,<}} \\ -\mu_{\mathcal{B}_{k,>}} \end{pmatrix} = B_k^T \begin{pmatrix} \lambda_{\mathcal{B}_{k,<}} \\ -\mu_{\mathcal{B}_{k,>}} \end{pmatrix} = -c_{\mathcal{B}_k}$$

$$\lambda_j(r_0 - AV_k y + \gamma_k 1_m)_j = 0, \quad \forall j \in \{1,\ldots,m\}$$
$$\mu_j(\gamma_k 1_m - (r_0 - AV_k y))_j = 0, \quad \forall j \in \{1,\ldots,m\}.$$

$$\Rightarrow \begin{matrix} \lambda_j = 0 \quad \forall j \in \mathcal{N} \\ \mu_j = 0 \quad \forall j \in \mathcal{N} \end{matrix}$$

Universiteit Antwerpen

Primal and Dual equation

Primal Equation

$$-\gamma_k = (r_0 - AV_k y_k)_i \qquad \forall i \in \mathcal{B}_{k,<}$$
$$(r_0 - AV_k y_k)_i = \gamma_k \qquad \forall i \in \mathcal{B}_{k,>}$$
$$|(r_0 - AV y_k)_i| \le \gamma_k \qquad \forall i \in \mathcal{N}_k$$

$$\Rightarrow \underbrace{\begin{pmatrix} -1_{\mathcal{B}_{k,<}} & AV_k|_{\mathcal{B}_{k,<}} \\ 1_{\mathcal{B}_{k,>}} & AV_k|_{\mathcal{B}_{k,>}} \end{pmatrix}}_{:=B_k} \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = B_k \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} = \begin{pmatrix} r_0|_{\mathcal{B}_{k,<}} \\ r_0|_{\mathcal{B}_{k,>}} \end{pmatrix}$$

Dual Equation and complementarity

$$1_m^T(\lambda + \mu) = 1,$$
$$V_k^T A^T(\lambda - \mu) = 0,$$
$$\lambda_j(r_0 - AV_k y + \gamma_k 1_m)_j = 0, \quad \forall j \in \{1, \ldots, m\}$$
$$\mu_j(\gamma_k 1_m - (r_0 - AV_k y))_j = 0, \quad \forall j \in \{1, \ldots, m\}.$$

$$\Rightarrow \begin{pmatrix} -1^T_{\mathcal{B}_{k,<}} & 1^T_{\mathcal{B}_{k,>}} \\ V_k^T A^T|_{\mathcal{B}_{k,<}} & V_k^T A|_{\mathcal{B}_{k,>}} \end{pmatrix} \begin{pmatrix} \lambda_{\mathcal{B}_{k,<}} \\ -\mu_{\mathcal{B}_{k,>}} \end{pmatrix} = B_k^T \begin{pmatrix} \lambda_{\mathcal{B}_{k,<}} \\ -\mu_{\mathcal{B}_{k,>}} \end{pmatrix} = -c_{\mathcal{B}_k}$$

$$\Rightarrow \begin{matrix} \lambda_j = 0 \quad \forall j \in \mathcal{N} \\ \mu_j = 0 \quad \forall j \in \mathcal{N} \end{matrix}$$

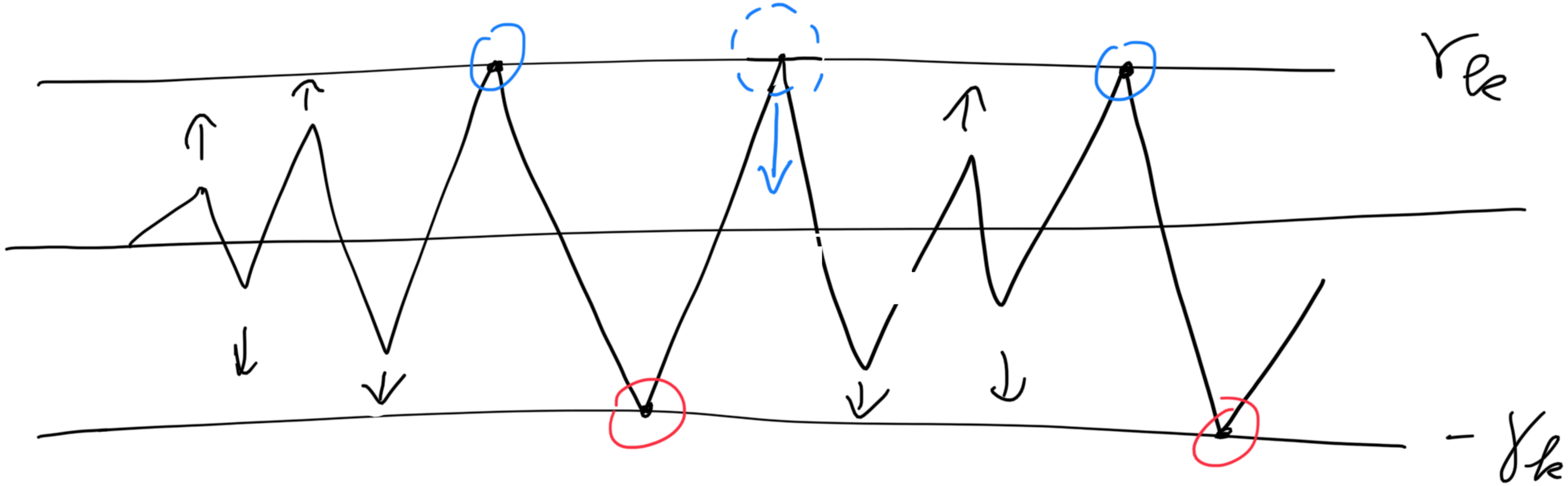
Universiteit Antwerpen

Pivot

If $\lambda \geq 0$ and $\mu \geq 0 \Rightarrow$ Optimal solution $x_k = x_0 + V_k y_k^*$ for $\mathcal{K}_k(A^T A, r_0)$

$\mathcal{B}_k^*$ is the optimal basic set

Else update the basic set with one index



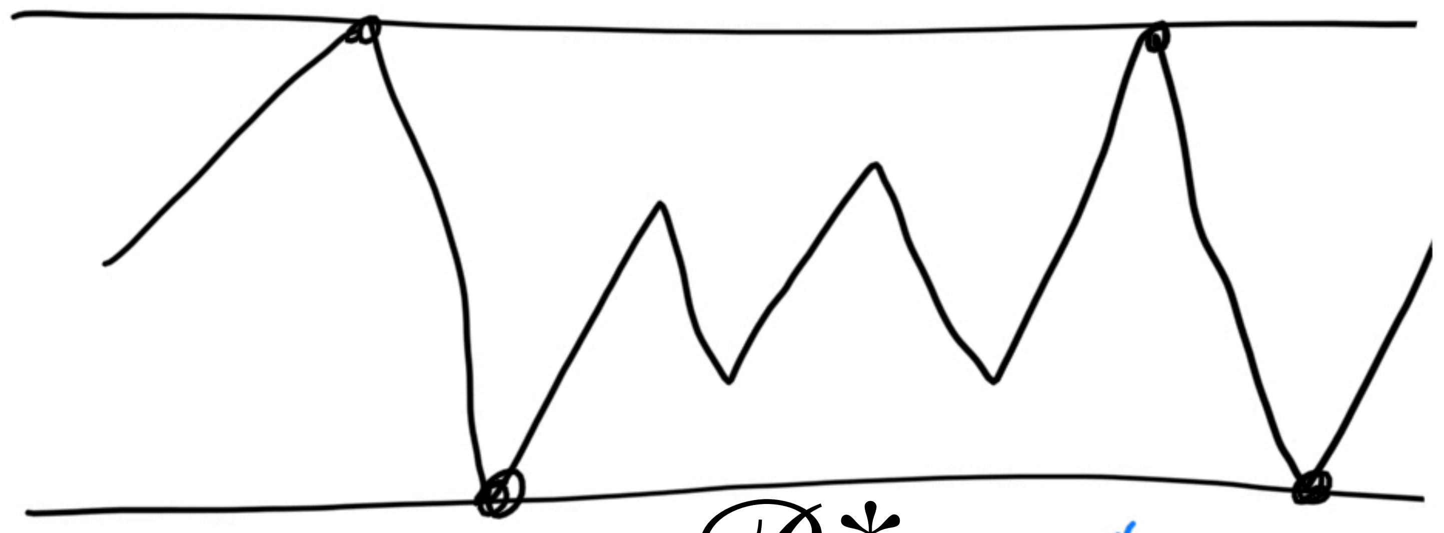
$$B_k \begin{pmatrix} \gamma_k^+ \\ y_k^+ \end{pmatrix} = B_k \begin{pmatrix} \gamma_k \\ y_k \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

$$\mathcal{B}_k' = \mathcal{B}_k \cup \{r\} \backslash \{q\}$$

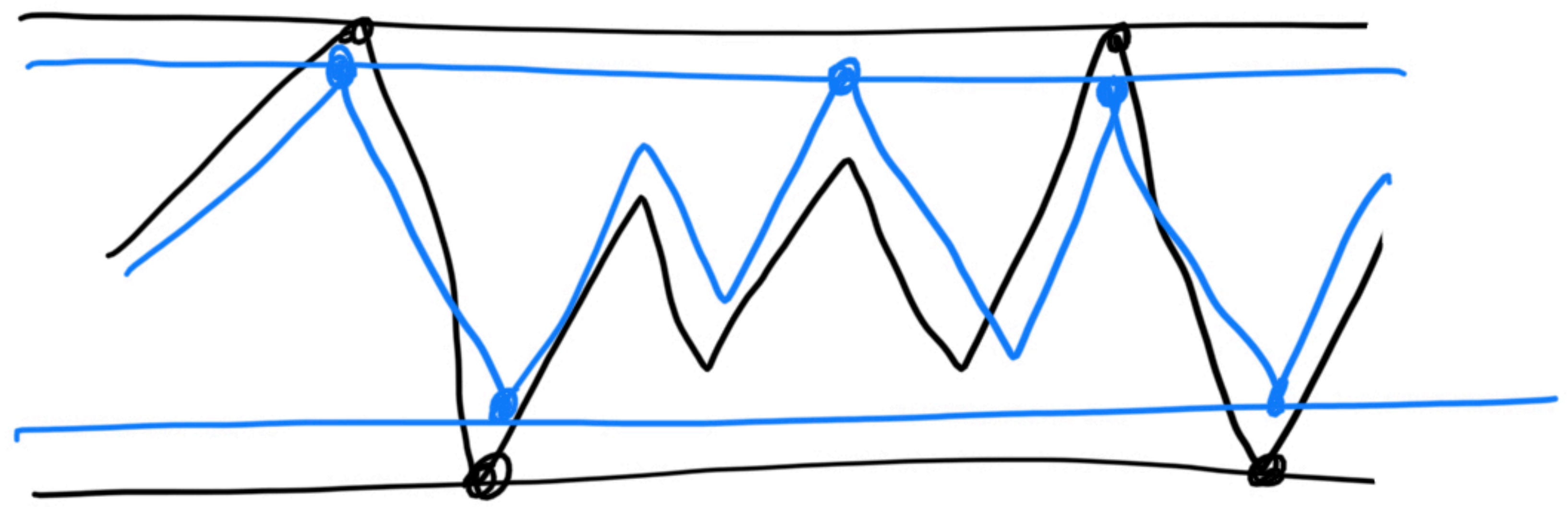
$$\mathcal{N}_k' = \mathcal{N}_k \cup \{q\} \backslash \{r\}$$

Expanding the Krylov subspace.

$$x_{k+1} = x_0 + \begin{pmatrix} V_k & v_{k+1} \end{pmatrix} y_{k+1} = x_0 + \begin{pmatrix} V_k & v_{k+1} \end{pmatrix} \begin{pmatrix} y_k^* + \Delta y_k \\ \alpha \end{pmatrix} = x_k + \begin{pmatrix} V_k & v_{k+1} \end{pmatrix} \begin{pmatrix} \Delta y_k \\ \alpha \end{pmatrix}$$



$$\mathscr{B}_k^*$$



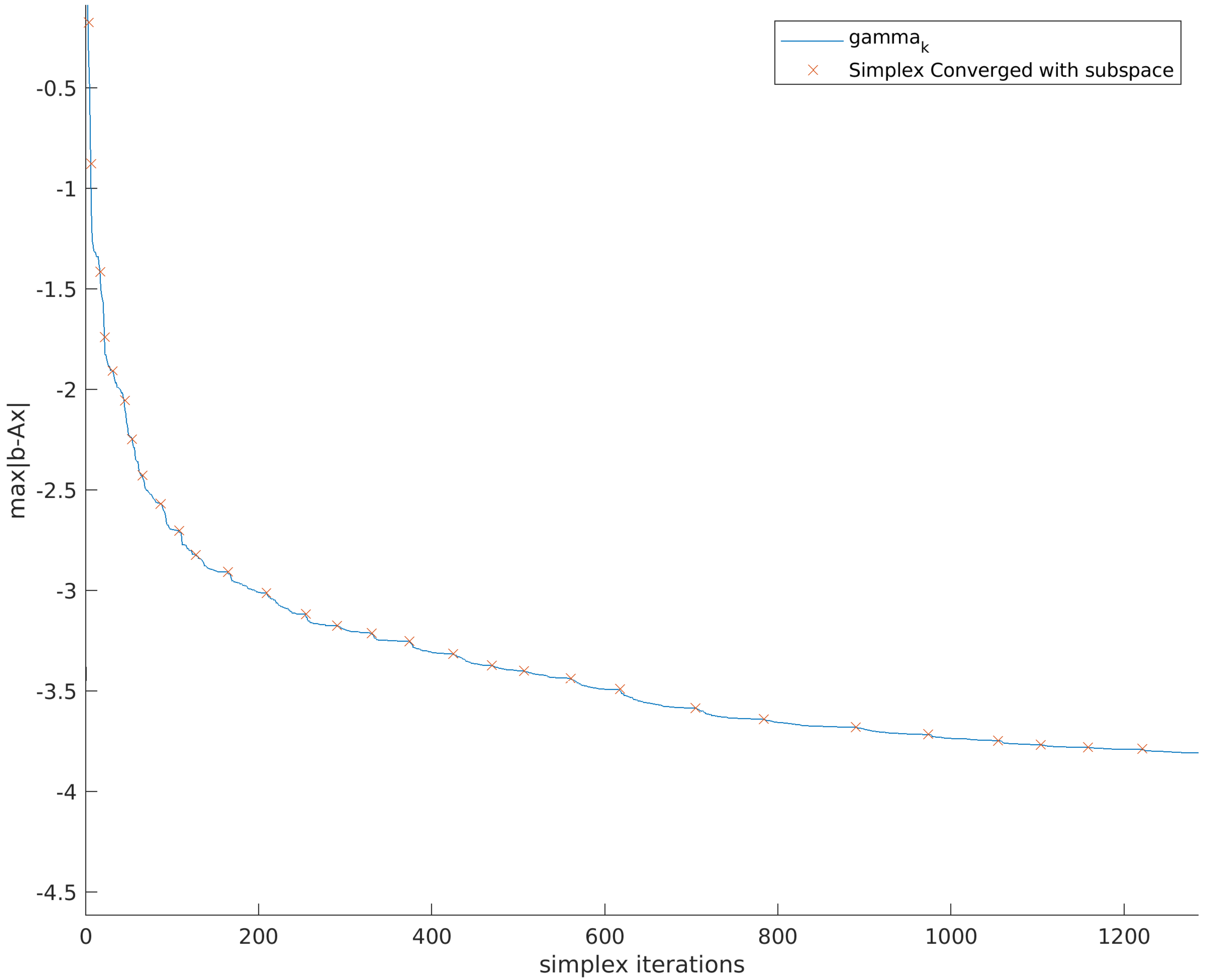
$$\mathscr{B}_{k+1} = \mathscr{B}_k^* \cup \{r\}$$

$$\begin{pmatrix} -1_{\mathscr{B}_{k,<}} & AV_k|_{\mathscr{B}_{k,<}} & Av_{k+1}|_{\mathscr{B}_{k,<}} \\ 1_{\mathscr{B}_{k,>}} & AV_k|_{\mathscr{B}_{k,>}} & Av_{k+1}|_{\mathscr{B}_{k,>}} \end{pmatrix} \begin{pmatrix} \gamma_k^* + \Delta\gamma_k \\ y_k^* + \Delta y_k \\ \alpha \end{pmatrix} = \begin{pmatrix} r_0|_{\mathscr{B}_{k,<}} \\ r_0|_{\mathscr{B}_{k,>}} \end{pmatrix}$$

$$y_{k+1} = \begin{pmatrix} y_k^* \\ 0 \end{pmatrix} + \begin{pmatrix} d_{1:k} \\ d_{k+1} \end{pmatrix} \Delta\gamma$$

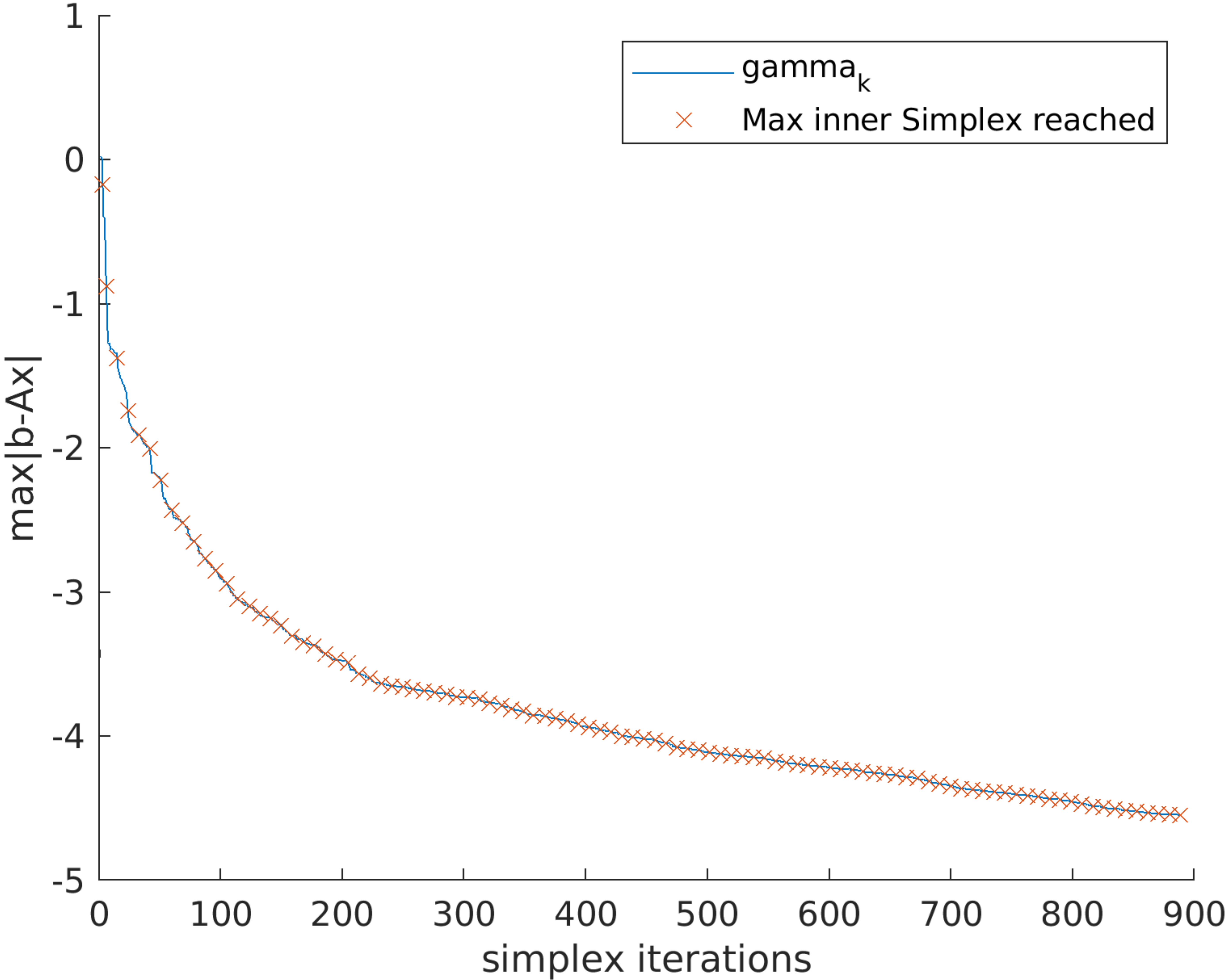
Universiteit Antwerpen

Convergence Krylov-Simplex for $\|b - Ax\|_\infty$



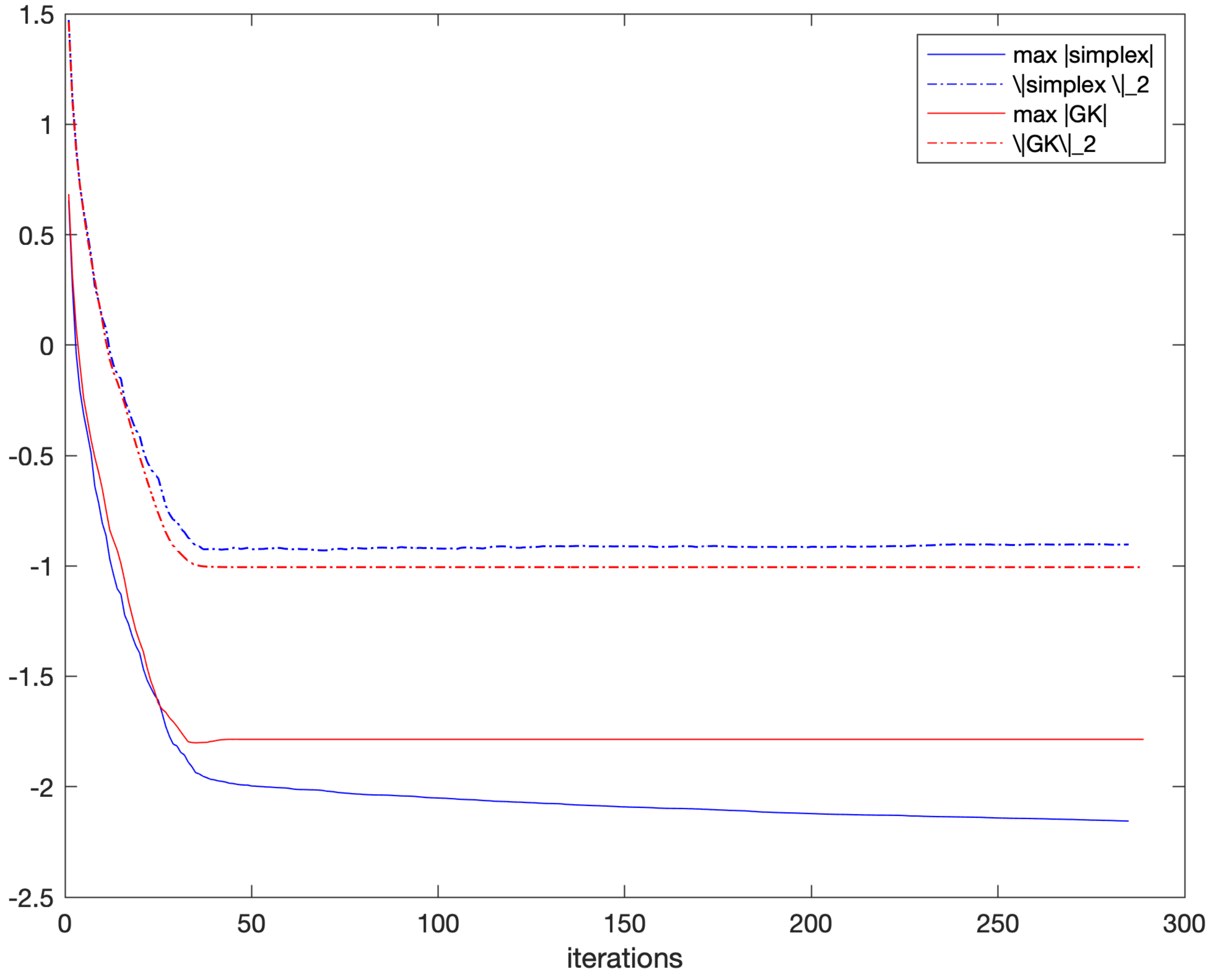
Universiteit Antwerpen

Convergence of Krylov-Simplex for $||b - Ax||_\infty$, limited inner iterations



Universiteit Antwerpen

Bounds for a Krylov-Simplex vs Golub-Kahan



$$\|b - Ax_{\text{ks}}\|_\infty \leq \|b - Ax_{\text{gk}}\|_\infty \leq \|b - Ax_{\text{gk}}\|_2 \leq \|b - Ax_{\text{ks}}\|_2$$

- Since $\|x\|_\infty \leq \|x\|_2$

Universiteit Antwerpen

Rank one updates to the basic matrix

Each iteration we have to solve the two $k \times k$ matrices:

$$B_k \binom{\gamma_k}{y} = \begin{pmatrix} r_0|_{\mathcal{B}_{k,<}} \\ r_0|_{\mathcal{B}_{k,>}} \end{pmatrix} \qquad B_k^T \begin{pmatrix} \lambda_{\mathcal{B}_{k,<}} \\ -\mu_{\mathcal{B}_{k,>}} \end{pmatrix} = -c_{\mathcal{B}_k}$$

Each inner simplex iteration we update the basic set. We do it in place.

$$\mathcal{B}_k' = \mathcal{B}_k \cup \{r\} \backslash \{q\}$$

So one column of B is substituted. We update its QR factorisation using QRUPDATE

$$B' = B + uv^T$$

$$Q'R' = QR + uv^T$$

Universiteit Antwerpen

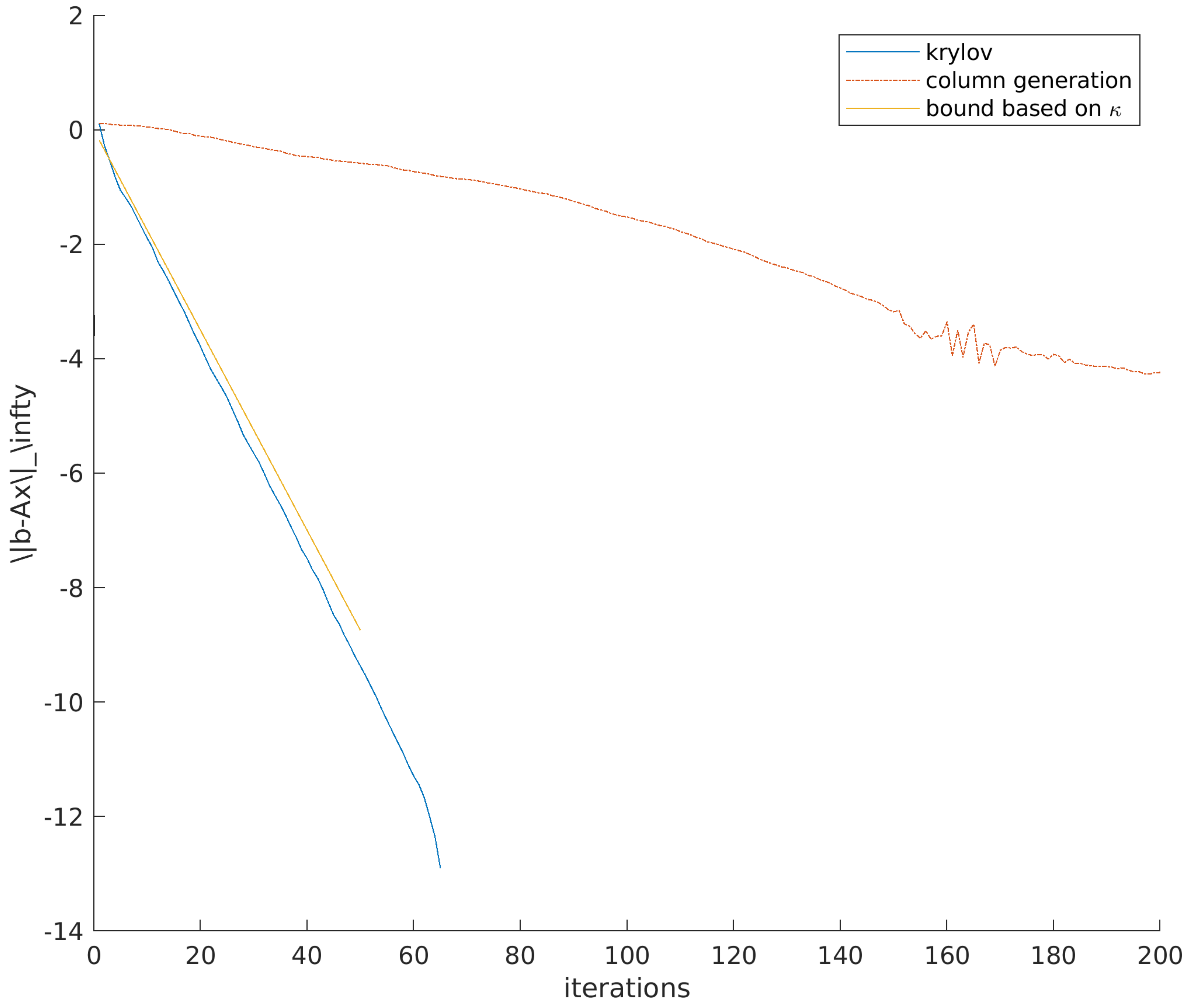
Expand the basis with lowest reduced cost.

$$\|r_{k+1}\|_\infty - \|r_k\|_\infty = \Delta\gamma_k = d_1 = [1\ 0\ \cdots\ 0]\, d$$
$$= -c^T_{\mathscr{B}^*_k} B_k^{-1} (Av_{k+1})_{\mathscr{B}^*_k} = (\lambda - \mu)^T A v_{k+1}$$
$$= c^T_{\text{reduced}} v_{k+1} \quad \text{with} \quad c_{\text{reduced}} = A^T(\lambda - \mu)$$

- Where λ and μ are the Lagrange multipliers from the optimal solution in the previous subspace.
- Choose $v_{k+1} = -A^T(\lambda - \mu)$ as the next vector to expand the basis.
- In *Column Generation* we solve minimal path with $c_{\text{reduced}} = (c - A^T\lambda)^T v_{k+1}$ with some additional constraints on the path *v*.

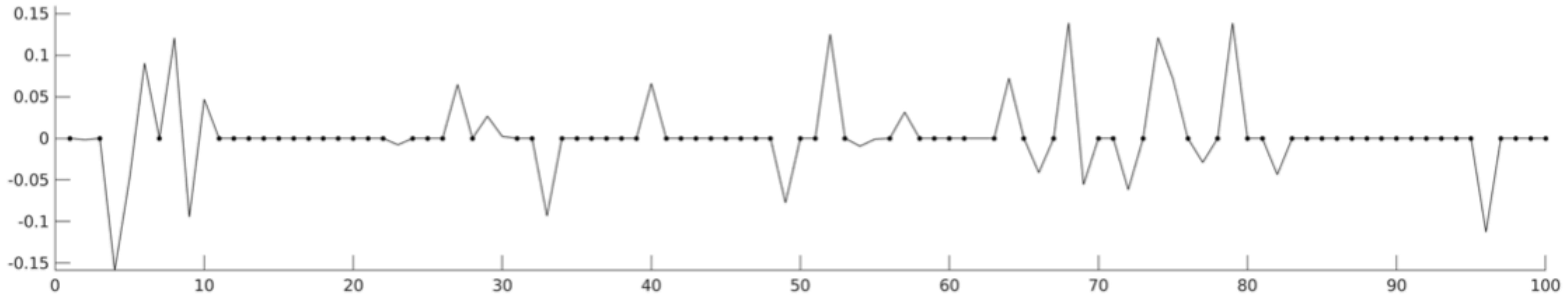
Universiteit Antwerpen

Krylov-Simplex vs Column Generation



Similar Krylov-Simplex methods for ℓ_1-norm

$$x_k := \text{argmin}_{x \in x_0 + \mathcal{K}(A^T A, r_0)} \|r_k\|_1$$



Vanroose, W., & Cornelis, J. (2021). Krylov-Simplex method that minimizes the residual in ℓ_1-norm or ℓ_∞-norm. *arXiv preprint arXiv:2101.11416*.

Universiteit Antwerpen

Summary Krylov-Simplex

- We minimise $||b - Ax||_\infty$ over a Krylov subspace.
  - Leads to a small LP problem with dense matrices that is solved with simplex method.
  - Krylov-Simplex: Krylov outer iteration, simplex inner iteration.
  - As the basis expands, the basic set gets one additional active constraint.
  - Early stopping of the inner iteration reduces total time to solution.
  - The factorisation of basic matrix is updated each iteration. We can converge using ‘qrupdates’ operations only.
- Expanding the basis with a column with smallest reduced cost not give improved convergence.

III: Updating the regularisation parameter while Krylov converges. Krylov-Newton.

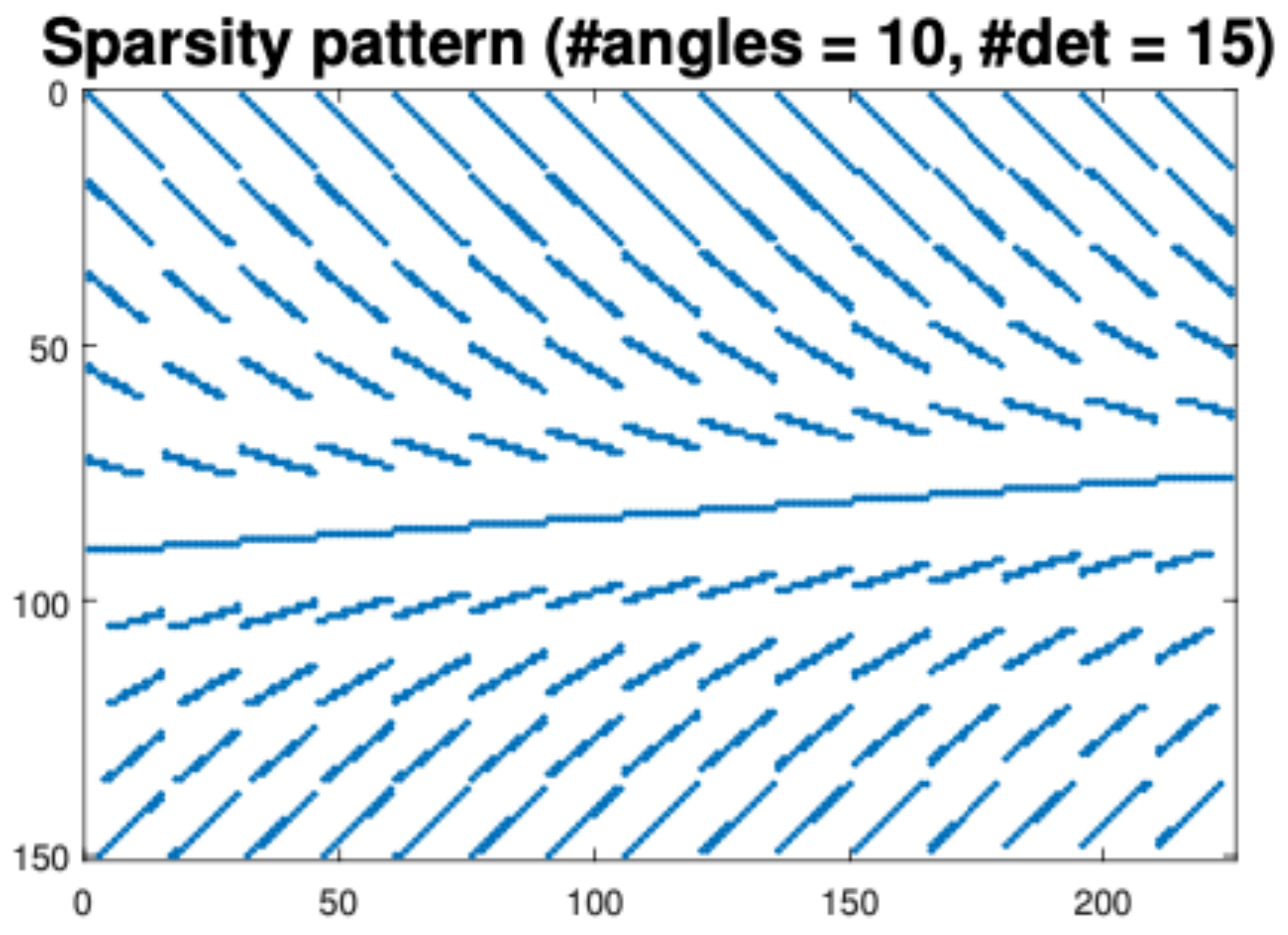
Universiteit Antwerpen

Computed Tomography example.

Consider a **Computed Tomography** example with 270 projection angles and 256 detectors (ASTRA toolbox):

$$Ax = b = b_{ex} + e = Ax_{ex} + e$$

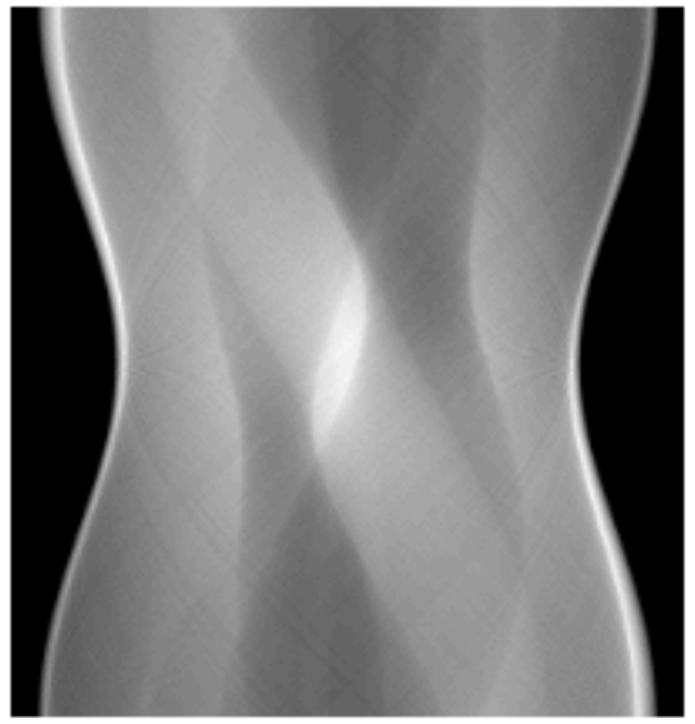
with $m = 69,120$ and $n = 65,536$.



x_{ex}

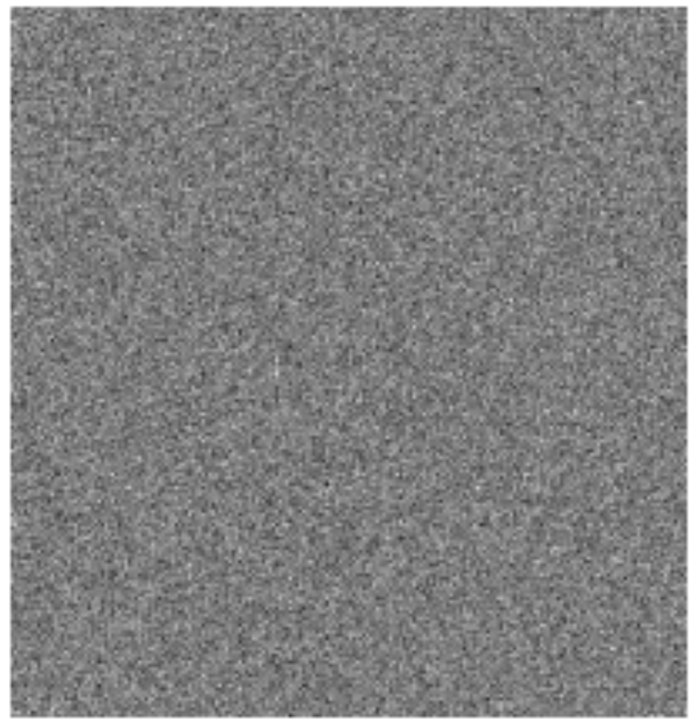
Exact image

$Ax_{ex} = b_{ex}$



Exact sinogram

e



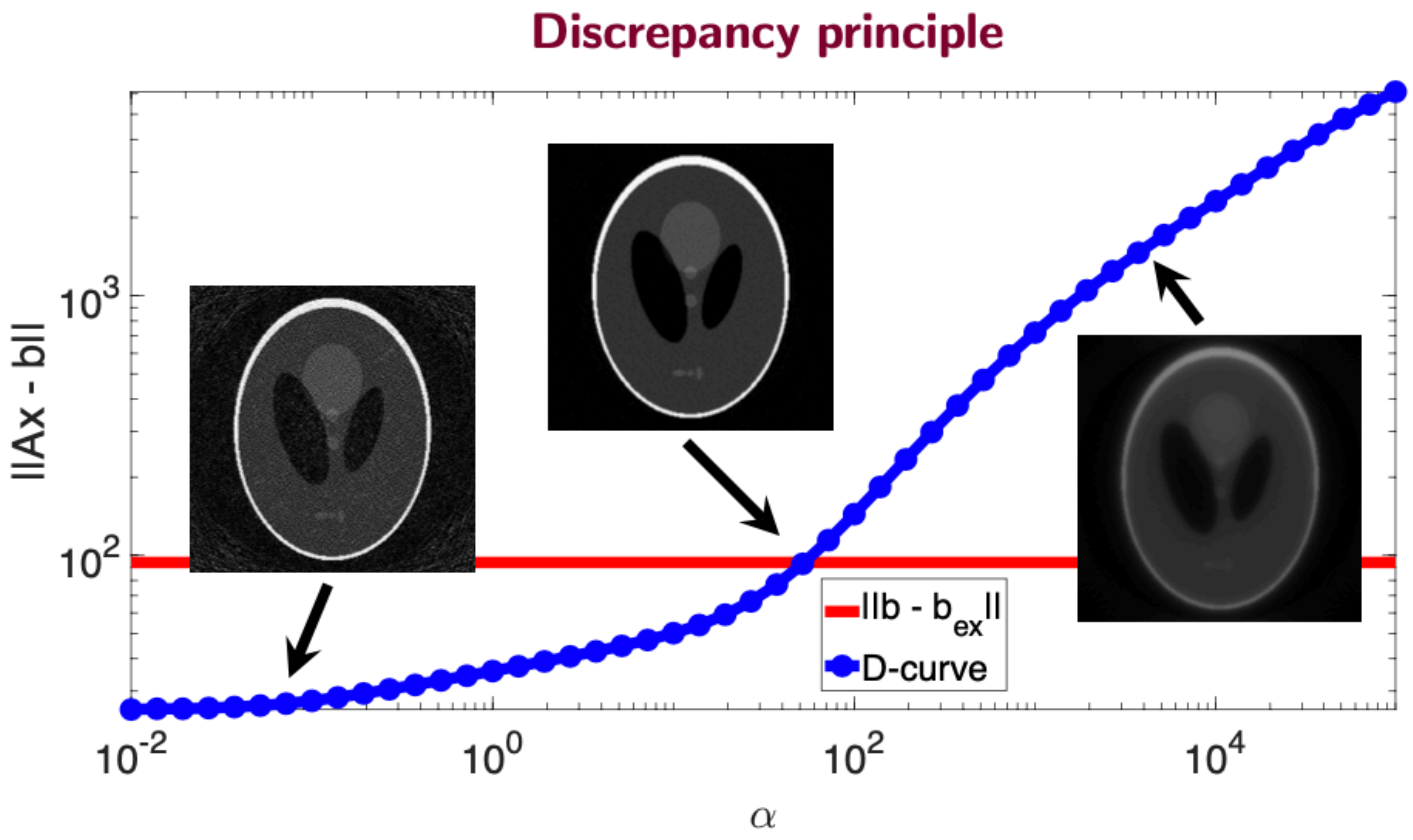
1% Gaussian noise

$\arg\min ||Ax - b||$

Reconstruction

Regularisation parameter and Discrepancy principle.

Compute $x_\alpha = \text{argmin}_{x \in \mathbb{R}^n} ||Ax - b||_2^2 + \alpha ||x||_2^2$ for many values of α



Universiteit Antwerpen

Discrepancy principle and Tikhonov regularisation

The constrained optimisation problem is

$$\min \frac{1}{2}\|x\|_2^2 \quad \text{subject to} \quad \|Ax - b\|_2^2 = \|e\|_2^2$$

This is a *Quadratic Constraint Quadratic Programming problem* (QCQP)

Has a unique solution with x^* with a Lagrange multiplier $\lambda > 0$. The KKT condition is

$$F(x,\lambda) = \begin{pmatrix} \lambda A(Ax - b) + x \\ \frac{1}{2}\|Ax - b\|_2^2 - \frac{1}{2}\|e\|^2 \end{pmatrix} = 0$$

This is a nonlinear system of equations.

The first equation is equivalent to Tikhonov regularised normal equation with a regularisation $\alpha = 1/\lambda$

$$\lambda A^T(Ax - b) + x = 0 \Leftrightarrow A^T(Ax - b) + \alpha x = 0$$

Universiteit Antwerpen

Classical Newton-Krylov Iteration

We use **Newton's** method to solve $F(x, \lambda) = 0$. We start with an initial guess (x_0, λ_0) with $\lambda_k > 0$ and updates

$$(x_{k+1}, \lambda_{k+1}) = (x_k, \lambda_k) + \gamma_k(\Delta x_{k+1}, \Delta\lambda_{k+1})$$

that are solutions of

$$J(x_k, \lambda_k)\begin{pmatrix}\Delta x_{k+1} \\ \Delta\lambda_{k+1}\end{pmatrix} = -F(x_k, \lambda_k)$$

The step size $\gamma_k \in [0,1[$ are adapted such that $\lambda_k + \Delta\lambda_{k+1} > 0$. The linear system is solved by a Krylov method, e.g MINRES.

Outer iteration is Newton, it linearises the system. Inner iteration is a Krylov subspace that is build each outer iteration and thrown away.

Bidiagonalisation

Since the matrix is non-square we use the bidiagonalization algorithm (**Bidiag1)** to generate matrices $V_k = [v_0, \ldots, v_{k-1}]$ and $U_{k+1} = [u_0, u_1, \ldots, u_k]$ with orthonormal columns that satisfy

$$AV_k = U_{k+1}B_{k+1}$$

With a lower bidiagonal matrix $B_{k+1} \in \mathbb{R}^{(k+1)\times k}$

$$\mathcal{R}(V_k) = \mathcal{K}_k(A^TA, A^Tb) = \text{span}\{A^Tb, (A^TA)A^Tb, \ldots, (A^TA)^{k-1}A^Tb\}$$

By writing $x_k = V_k y_k$ for $y_k \in \mathbb{R}^k$ we have that $\|x_k\| = \|V_k y\| = \|y_k\|$

$$\begin{aligned}\|Ax - b\| &= \|AV_k y_k - U_{k+1}c_k\| \\ &= \|U_{k+1}(B_{k+1}y_k - c_k)\| \\ &= \|B_{k+1}y_k - c_k\|\end{aligned}$$

With $c_k = \|b\|e_1$

Universiteit Antwerpen

Krylov-Newton

The projected minimisation problem is

$$\min_{x_k \in \mathcal{R}(V_k)} \|x_k\|_2^2 \quad \text{s.t.} \quad \|Ax - b\|_2^2 = \|e\|_2^2$$

is equivalent

$$\min_{y_k \in \mathbb{R}^k} \|y_k\|_2^2 \quad \text{s.t.} \quad \|B_{k+1}y_k - c_k\|_2^2 = \|e\|_2^2$$

It has KKT conditions

$$F^{(k)}(y,\lambda) = \begin{pmatrix} \lambda B_{k+1}(B_{k+1}y - c_k) + y \\ \frac{1}{2}\|B_{k+1}y - c_k\|^2 - \frac{1}{2}\|e\|^2 \end{pmatrix} = 0$$

A solution of $F^{(k)}(y,\lambda) = 0$ gives an approximate solution of $F(x,\lambda)$ but is much cheaper to solve if $k \ll n$

Universiteit Antwerpen

A Single Newton step gives a descent direction

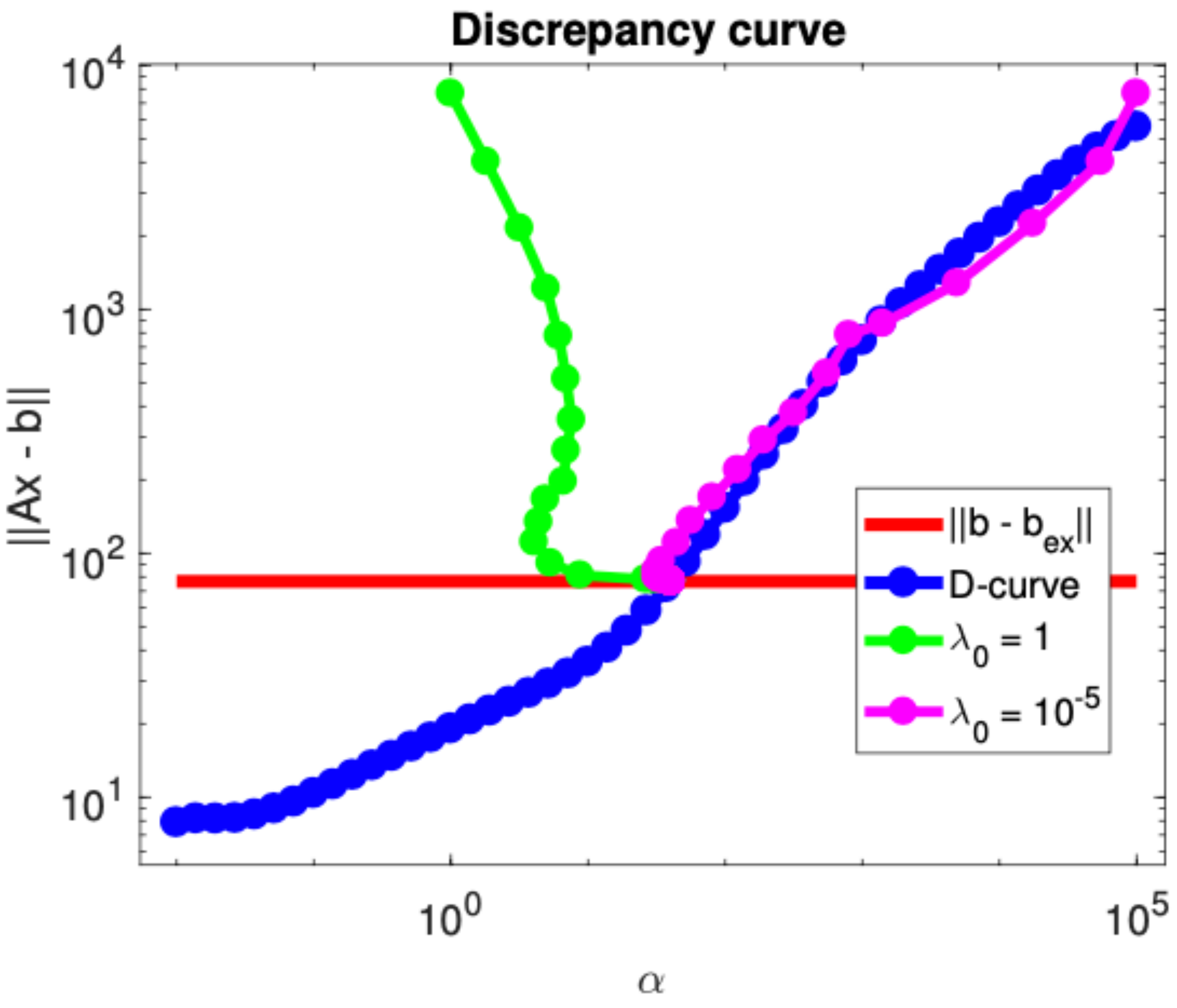
Theorem. *Let* $(\Delta y_k, \Delta\lambda_k)$ *be the Newton step for* $F^{(k)}(\bar{y}_k, \lambda_k)$ *and* $\Delta x = V_k \Delta y_k$ *then*

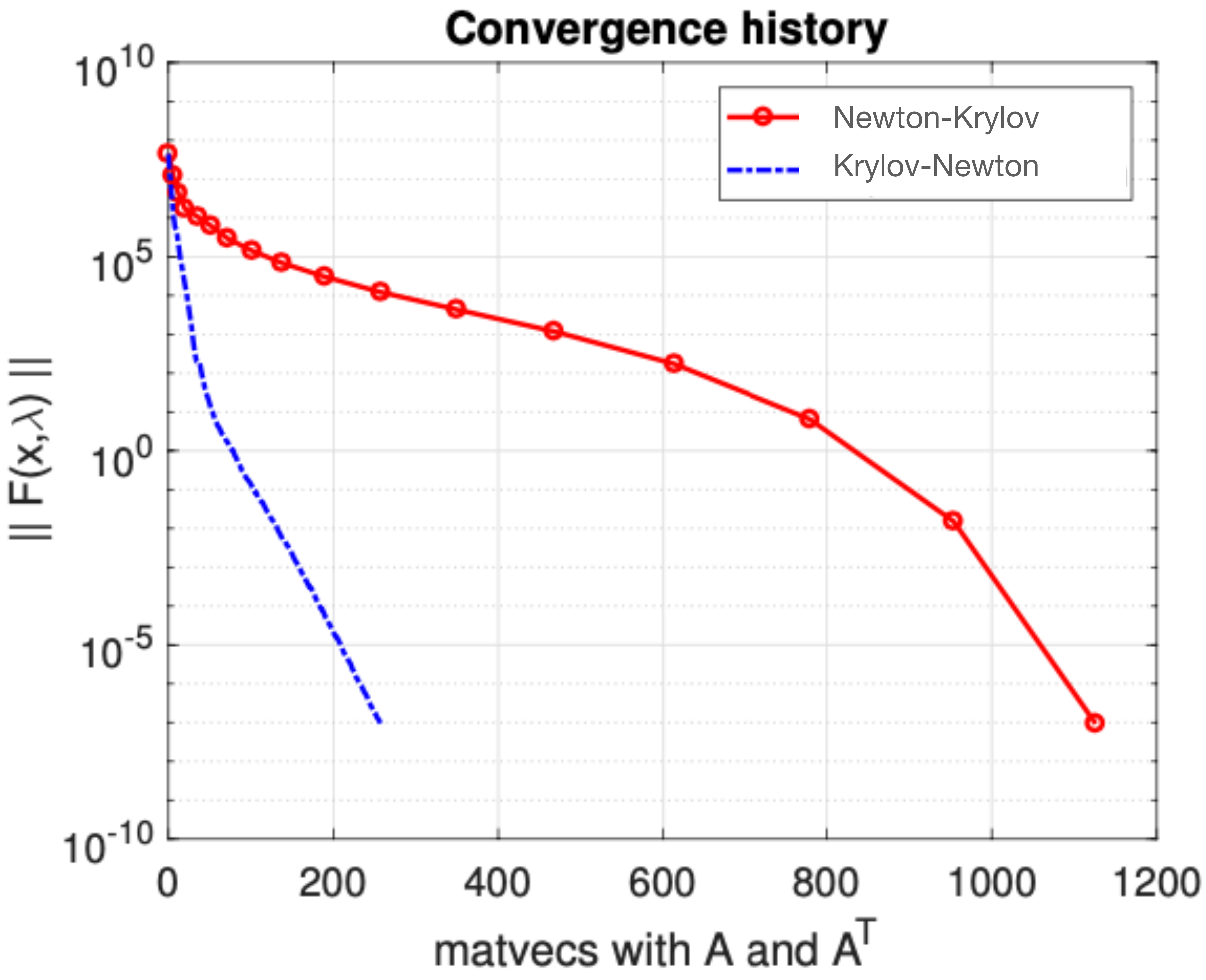
$$\Delta f(x_k, \lambda_k)^T \begin{pmatrix} \Delta x_{k+1} \\ \Delta\lambda_k \end{pmatrix} = -\|F(x_k, \lambda_k)\|^2 \leq 0$$

with $f(x, \lambda) = \dfrac{1}{2}\|F(x, \lambda)\|$

So the Newton direction is a descent direction for backtracking line search we then select a step size γ_k

$$\frac{1}{2}\|F(x_{k+1}, \lambda_{k+1})\| \leq \left(\frac{1}{2} - \rho\gamma_k\right)\|F(x_k, \lambda_k)\|^2$$





Universiteit Antwerpen

General Regularisation terms.

$$\min_{x \in \mathbb{R}^n} \quad \Psi(x) \quad \text{s.t.} \quad ||Ax - b||^2 = ||e||^2$$

with convex twice differentiable regularization function $\Psi(x) : \mathbb{R}^n \to \mathbb{R}$.

First order optimality conditions are given by $F(x, \lambda) = 0$ with

$$F(x, \lambda) = \begin{pmatrix} \lambda A^T(Ax - b) + \nabla\Psi(x) \\ \frac{1}{2}||Ax - b||^2 - \frac{||e||^2}{2} \end{pmatrix}.$$

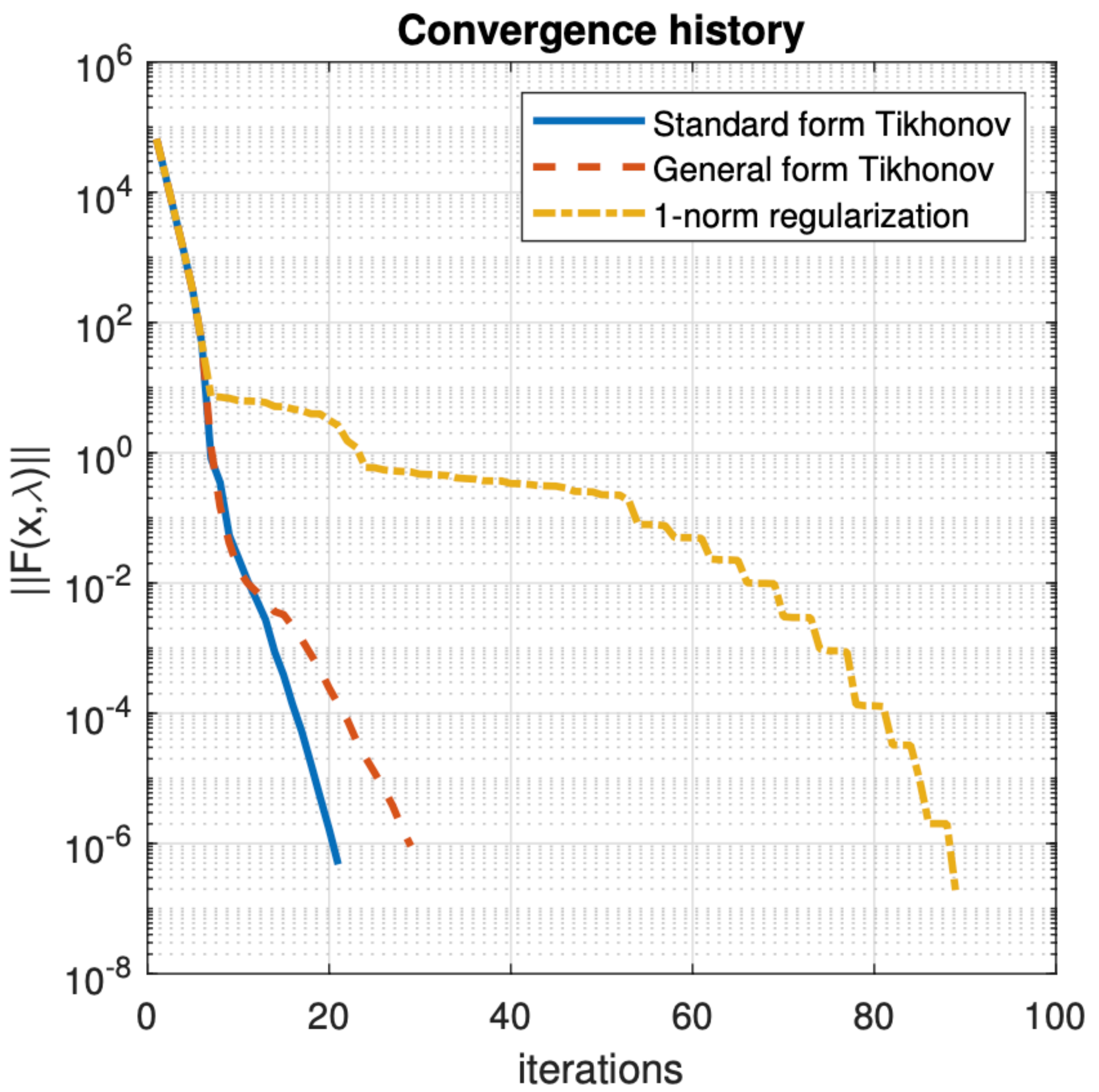
We now project the optimization problem on a **Generalized Krylov subspace**

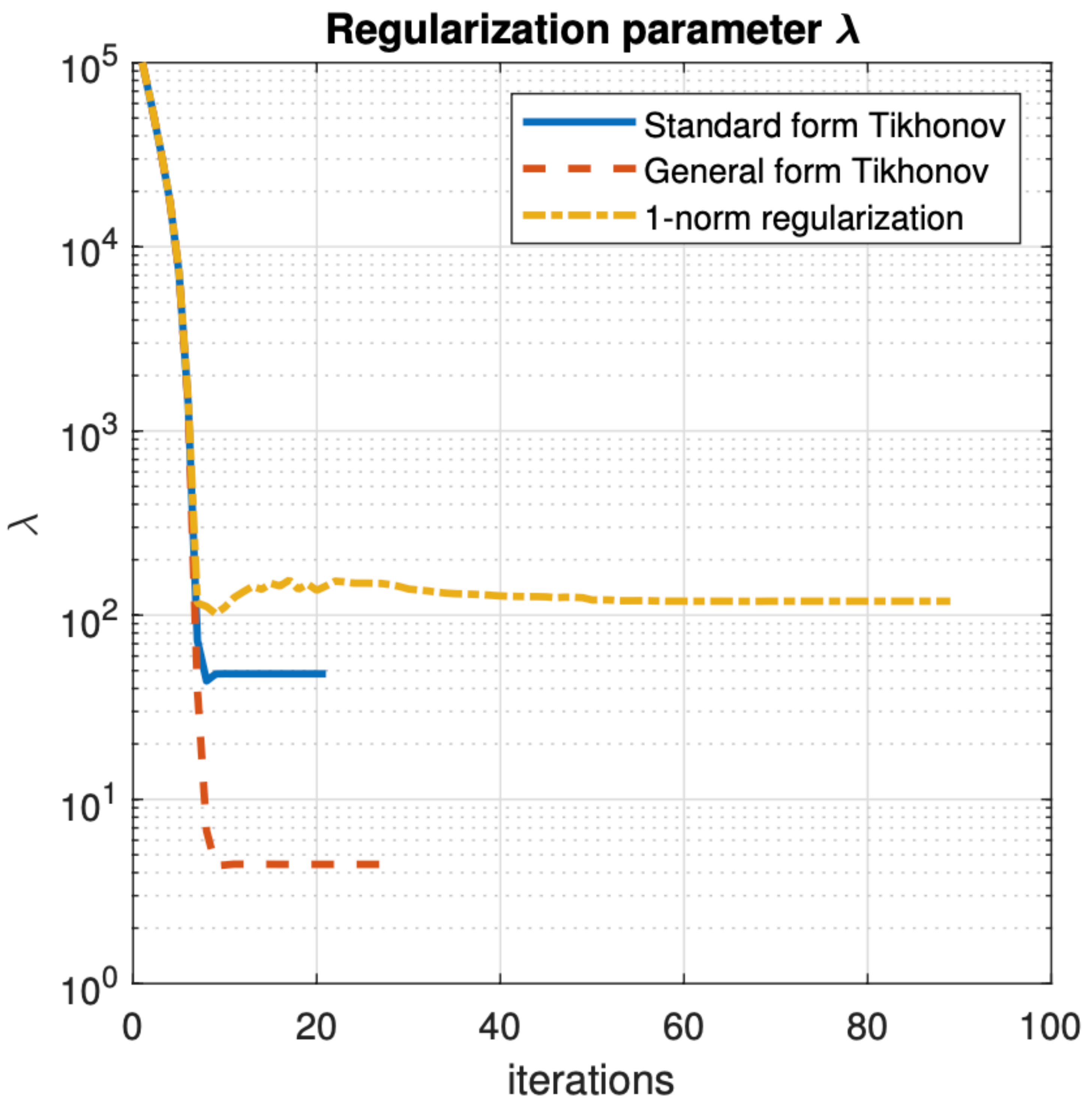
$$\mathcal{R}(V_k) = \text{span}\{r_0, r_1, \ldots, r_{k-1}\}$$

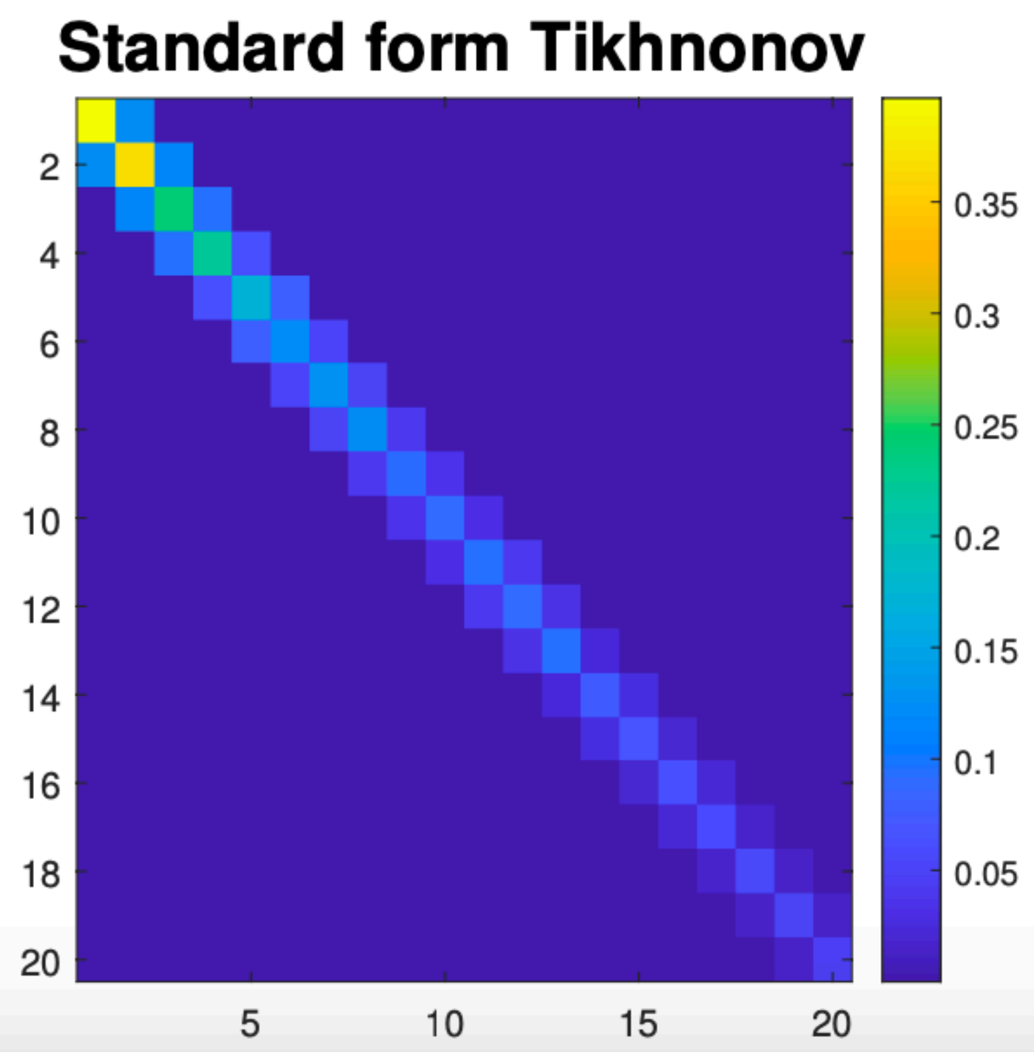
with orthonormal columns computed using Gram–Schmidt and

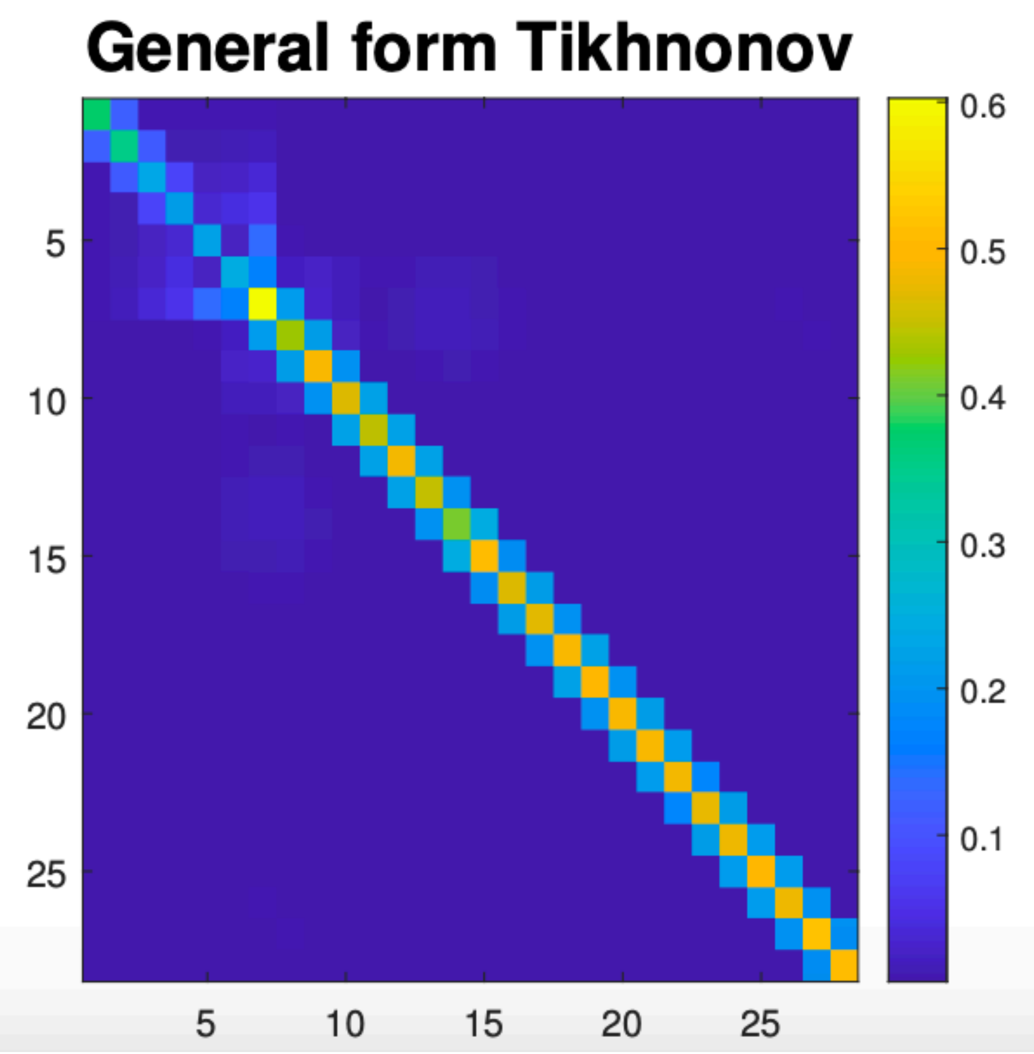
$$r_{k-1} = \lambda_{k-1} A^T(Ax_{k-1} - b) + \nabla\Psi(x_{k-1}).$$

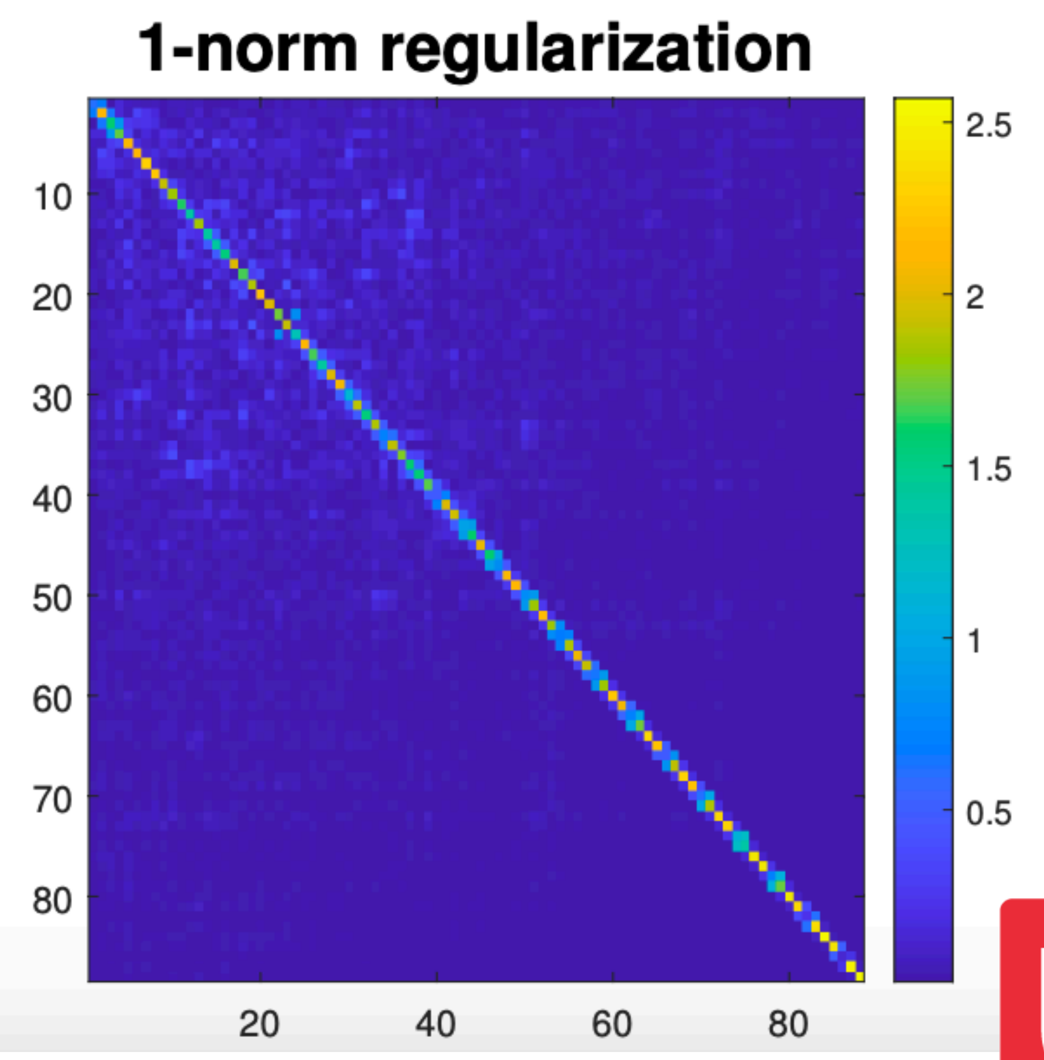
Universiteit Antwerpen











Universiteit Antwerpen

Summary Krylov-Newton

- We minimise $||Ax - b||^2 + \alpha||Lx||^2$ over a subspace but simultaneously adapt the regularisation parameter such that discrepancy principle is satisfied
  - Bases is expanded based on the residuals.
  - A single inner step is sufficient.
  - As soon a good guess for the Lagrange multipliers is found, we achieve Krylov convergence.
  - Krylov-Newton: <u>Krylov outer</u> iteration, Newton <u>inner</u> iteration

Cornelis J, Schenkels N, Vanroose W. *Projected Newton method for noise constrained Tikhonov regularization.* Inverse Problems. 2020 Apr 8;36(5):055002.
Cornelis J, Vanroose W. *Projected Newton method for noise constrained ℓ p regularization.* Inverse Problems. 2020 Dec 3;36(12):125004.

V: Krylov-Active Set. Regularisation by $l \leq x \leq u$

$$\min \|Ax - b\|_2^2$$

$$\text{s.t. } l \leq x \leq u$$

Problem formulation

Problem

$$\min \|Ax - b\|_2^2$$
$$\text{s.t. } l \leq x \leq u \quad \Rightarrow$$

KKT conditions

$$A^T(Ax - b) - \lambda + \mu = 0$$
$$\lambda_i(x_i - \ell_i) = 0 \qquad i \in \{1, \ldots, m\}$$
$$\mu_i(u_i - x_i) = 0 \qquad i \in \{1, \ldots, m\}$$
$$\ell_i \leq x_i \leq u_i \qquad i \in \{1, \ldots, m\}$$
$$(\lambda, \mu) \geq 0$$

$$\min \|AV_k y_k - b\|_2^2$$
$$\text{s.t. } l \leq V_k y \leq u \quad \Rightarrow$$

$$V_k^T A^T(AV_k y_k - b) - V_k^T \lambda + V_k^T \mu = 0$$
$$\lambda_i([V_k y_k]_i - \ell_i) = 0 \qquad i \in \{1, \ldots, m\}$$
$$\mu_i(u_i - [V_k y_k]_i) = 0 \qquad i \in \{1, \ldots, m\}$$
$$\ell_i \leq [V_k y_k]_i \leq u_i \qquad i \in \{1, \ldots, m\}$$
$$(\lambda, \mu) \geq 0$$

Universiteit Antwerpen

Optimality in subspace leads to orthogonality of residuals.

$$r_k = A^T(AV_ky_k - b) - \lambda + \mu$$

From KKT conditions we have

$$\begin{aligned} 0 &= V_k^T A^T(AV_ky_k - b) - V_k^T\lambda + V_k^T\mu \\ &= V_k^T(A^T(AV_ky_k - b) - \lambda + \mu) \\ &= V_k^T r_k \\ &\Rightarrow r_k \perp V_k \end{aligned}$$

We use as subspace basis

$$V_k = [r_0, r_1, r_2, \ldots,]$$

Universiteit Antwerpen

Algorithm 1 Residual QPAS subspace (ResQPASS)

Require: $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m,\ l, u \in \mathbb{R}^n$

$r_0 = A^T b, V_1 = [r_0/\|r_0\|], y_1 = 0, \mathcal{W}_1 = \emptyset$

for $k = 1, 2, \dots, m$ **do**

  $y_k^*, \lambda_k, \mu_k, \mathcal{W}_k^* \leftarrow$ Solve equation (1) using QPAS, guess y_k and working set $\mathcal{W}_k$

  $r_k = A^T(AV_k y_k^* - b) - \lambda_k + \mu_k$

  if $\|r_k\|_2 \leq tol$ **then**

    $x = V_k y_k$ break;

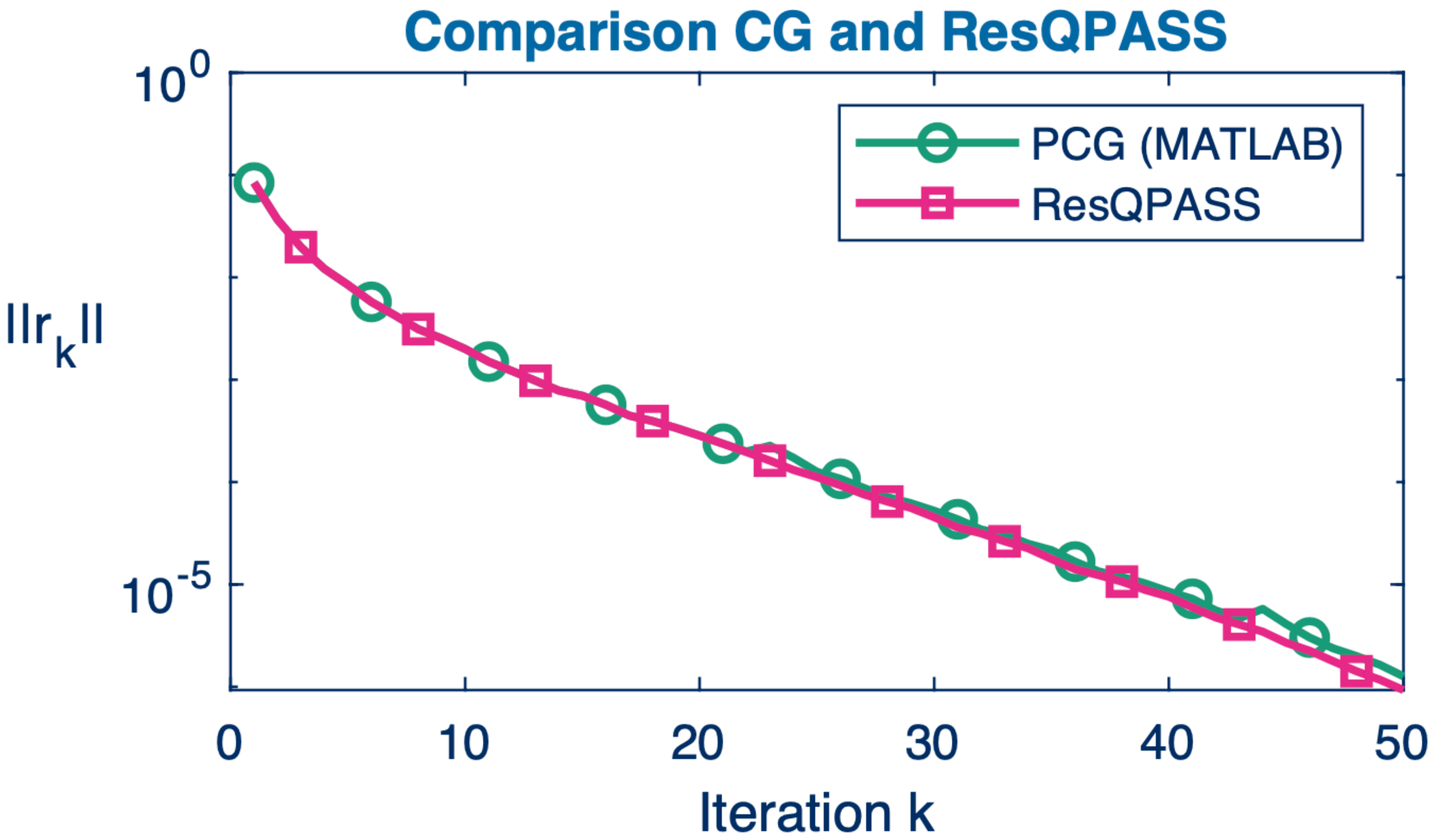
  end if

  $V_{k+1} \leftarrow \begin{bmatrix} V_k & r_k/\|r_k\| \end{bmatrix}$

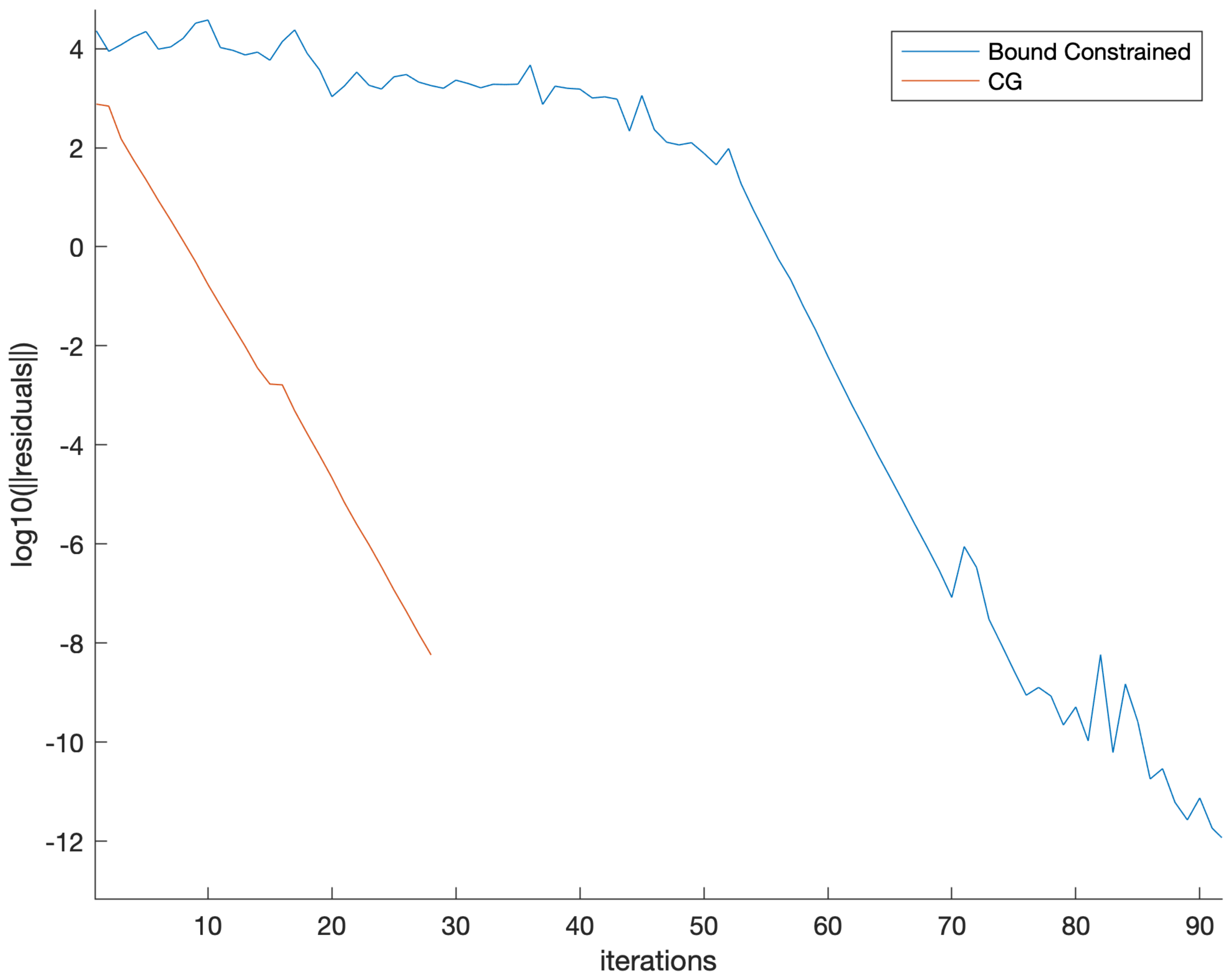
  $y_{k+1} \leftarrow \begin{bmatrix} (y_k^*)^T & 0 \end{bmatrix}^T$

  $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k^*$

end for


Comparison CG and ResQPASS
10^0
10^{-5}
$||r_k||$
PCG (MATLAB)
ResQPASS
0
10
20
30
40
50
Iteration k

A few active constraints.



Universiteit Antwerpen

Why Krylov convergence?

$$\min \|AV_k y_k - b\|_2^2$$

s.t. $l \leq V_k y \leq u$

$$\operatorname{span}\left\{-\lambda_1 + \mu_1, -\lambda_2 + \mu_2, \ldots, -\lambda_k + \mu_k\right\},$$

$$r_k = A^T(AV_k y_k - b) - \lambda + \mu$$

$$\operatorname{span}\left\{r_1 - A^T(AV_1 y_1 - b), r_2 - A^T(AV_2 y_2 - b), \ldots, r_k - A^T(AV_k y_k - b)\right\}.$$

$$-\lambda_k + \mu_k = r_k - A^T AV_k y_k = \sum_{l=1}^{k-1} (r_l - A^T(AV_l y_l - b))a_l \in \mathcal{R}[V_k, A^T AV_{k-1}],$$

Why Krylov convergence?

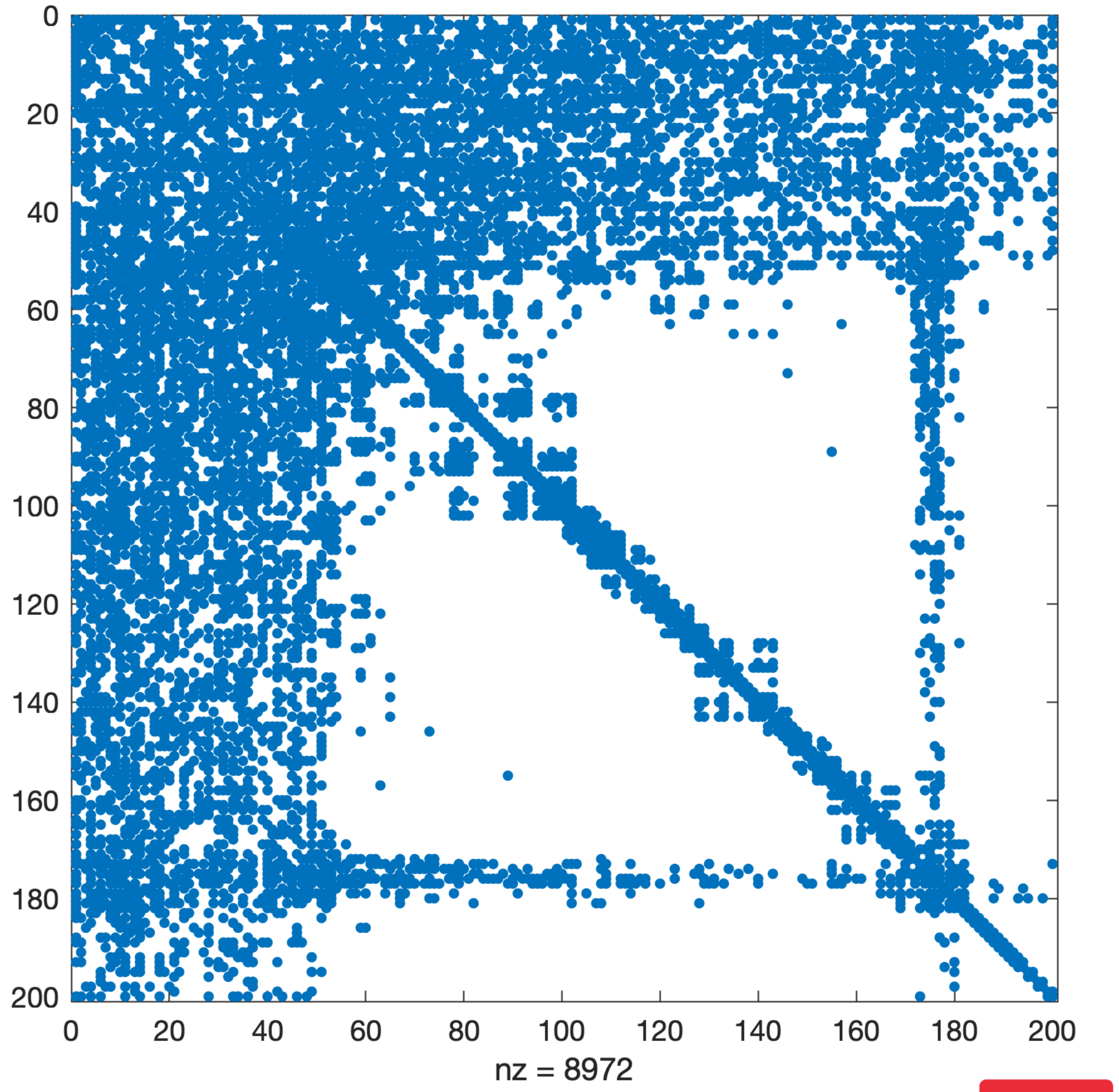
$$\begin{aligned} r_k &= A^T(AV_k y_k - b) - \lambda_k + \mu_k \\ &= A^T A V_k y_k + V_{k_0}\alpha^{(k)} + A^T A V_{k_0-1}\beta^{(k)} \\ &= A^T A V_k (y_k + (\beta^{(k)}, 0)^T) + V_{k_0}\alpha^{(k)} \quad \forall k \geq k_0. \end{aligned}$$

$$r_k = A^T(AV_k y_k - b) - \lambda_k + \mu_k = \sum_{m=0}^{k-k_0+1} (A^T A)^m V_{k_0} \gamma_m^{(k)} \quad \forall k \geq k_0.$$

Bas Symoens & Wim Vanroose, *Residual QPAS subspace (ResQPASS) algorithm for bounded-variable least squares (BVLS) with superlinear Krylov convergence*: Arxiv: 2302.13616

Asymptotic Krylov Structure.

$$V_k^T A^T A V_k$$



Bas Symoens. *Deelruimtemethode voor Inverse problemen,* Masterscriptie 2022.
Bas Symoens & Wim Vanroose, *Residual QPAS subspace (ResQPASS) algorithm for bounded-variable least squares (BVLS) with superlinear Krylov convergence:* Arxiv: 2302.13616

Universiteit Antwerpen

Conclusions.

- Expand the subspace as the *outer iteration*. Solve a small projected LP, QP or non-linear system with an *inner iteration*.
    - *Krylov-Simplex*. Outer iteration Krylov, inner simplex iterations.
    - *Krylov-Newton*: Outer iteration expands using the residuals, inner iteration updates regularisation parameter. Asymptotically a classical Krylov subspace.
    - *Krylov-Active Set*. Outer iteration expands using the residuals. As soon as the bounds are discovered, classical Krylov convergence sets in.
- Future work:
    - Use interior point methods as for inner problem.
    - Accelerate column generation.
    - Accelerate general LP and QP problems beyond inverse problems?
    - Can we combine with branching and cutting?

Universiteit Antwerpen