Convergence of Stochastic Iterative Algorithms for Singular Systems

Mengdi Wang and Dimitri P. Bertsekas Laboratory of Information and Decision Systems, MIT {mdwang, dimitrib} @mit.edu

The Problem

- Want to solve a **singular** linear system:

$$Ax = b,$$

based on approximations A_k, b_k such that

$$A_k \to A, \qquad b_k \to b.$$

- Focus on iterative algorithm (since $A_k^{-1} b_k$ does not work):

$$x_{k+1} = x_k - \gamma G_k (A_k x_k - b_k).$$

- Related classical algorithms with different choices of G_k:
  - ◇ Projection/proximal/splitting algorithms.

Deterministic Convergence

Theorem 1 *The algorithm*

$$x_{k+1} = x_k - \gamma G(Ax_k - b)$$

converges to a solution of $Ax = b$ for sufficiently small $\gamma > 0$ if and only if the following conditions hold:

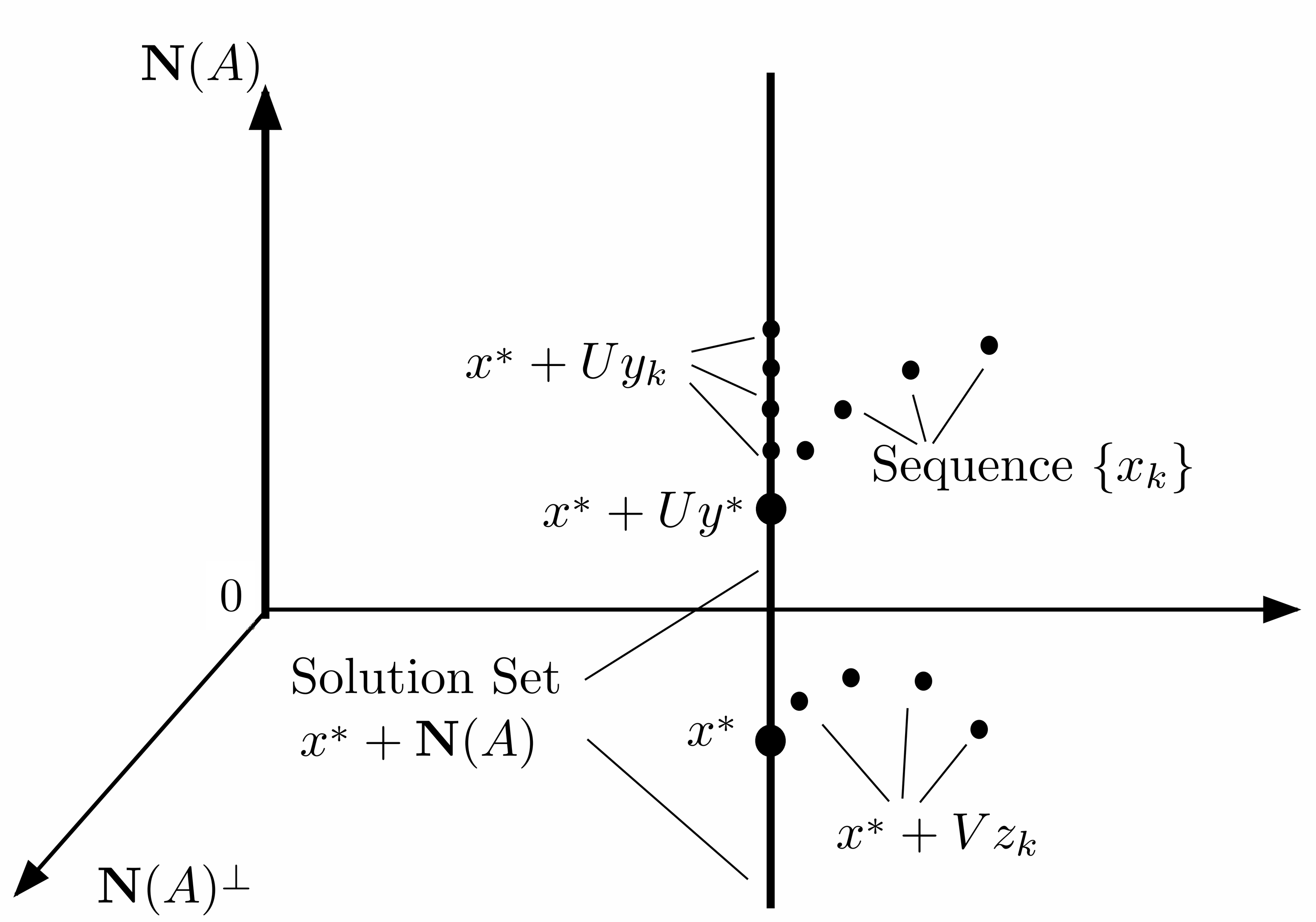
(a) Each eigenvalue of GA either has a positive real part or is 0.

(b) $\dim(\mathbf{N}(GA))$ = algebraic multiplicity of eigenvalue 0 of GA.

(c) $\mathbf{N}(A) = \mathbf{N}(GA)$.

Iteration Decomposition

- Illustration of the convergence process of the deterministic iteration.



- The iteration decomposes into two orthogonal components, on $\mathbf{N}(A)$ and $\mathbf{N}(A)^\perp$, respectively.
- $x_k = x^* + Uy_k + Vz_k$.

Stochastic Simulation

- Sample Average Approximation:
  - ◇ Generate an infinite sequence of random variables

$$\{(W_t, v_t) \mid t = 1, 2, \ldots\}.$$

  - ◇ Estimates A and b with A_k and b_k given by

$$A_k = \frac{1}{k}\sum_{t=1}^{k} f(W_t), \qquad b_k = \frac{1}{k}\sum_{t=1}^{k} g(v_t).$$

  - ◇ The simulation error sequence satisfies

$$E_k = (A_k - A, G_k - G, b_k - b) \xrightarrow{a.s.} 0.$$

- **Stochasticity + Singularity → Divergence.**

Theorem 2 *There exist examples of proximal/projection methods with convergent deterministic counterparts such that when using sample average approximation:*

$$\limsup_{k\to\infty} \|x_k\| = \infty, \qquad \limsup_{k\to\infty} \|Ax_k - b\| = \infty, \qquad w.p.1.$$

Stabilizing the Stochastic Algorithms

- Convergent stabilized algorithms (with $\delta_k \downarrow 0$ *slowly*):
  - ◇ Shift eigenvalues of $I - \gamma G_k A_k$ by $-\delta_k$:

$$x_{k+1} = (1 - \delta_k)x_k - \gamma G_k (A_k x_k - b_k).$$

  - ◇ Other stabilization methods exist.

A General Convergence Theorem of Stabilization

- A general form of stabilized algorithm:

$$x_{k+1} = T_k x_k + g_k,$$

where T_k and g_k are approximations of T and g:

$$T_k = T + \delta_k D + O(\delta_k^2 + \|E_k\|), \qquad g_k = g + \delta_k d + O(\delta_k^2 + \|E_k\|).$$

Theorem 3 *Let $x_{k+1} = Tx_k + g$ be convergent. Assume that there exists an invertible matrix P such that*

$$\|T_k\|_P \le 1 - O(\delta_k), \qquad \forall\ k \text{ sufficiently large},$$

and let $\{\delta_k\}$ satisfy $E_k/\delta_k \xrightarrow{a.s.} 0$ and $\sum_{k=0}^{\infty} \delta_k = \infty$. Then for all x_0, the algorithms $x_{k+1} = T_k x_k + g_k$ is such that

$$x_k \xrightarrow{a.s.} \hat{x},$$

where $\hat{x}$ is the unique solution to the system $\hat{\Pi}(Dx + d) = 0, x = Tx + g$, with $\hat{\Pi}$ the projection onto $\mathbf{N}(A)$ w.r.t. $\|\cdot\|_P$.

Algorithms with Naturally Convergent Residuals

- Nullspace-Consistent Iteration
  - ◇ Example: Projected equations in approximate DP.
- Regularized Regression
  - ◇ Proximal iteration applied to $A'\Sigma^{-1}Ax = A'\Sigma^{-1}b$.
  - ◇ Naturally contractive structure, resilient to simulation noise.

Residual Convergence Theorems

Theorem 4 (Nullspace-Consistent Iteration) *If its deterministic counterpart converges and it is nullspace-consistent, i.e.,*

$$\mathbf{N}(G_k A_k) = \mathbf{N}(A), \textit{ eventually},$$

the iteration $x_{k+1} = x_k - \gamma(G_k A_k - b_k)$ generates $\{x_k\}$ such that $Ax_k - b \xrightarrow{a.s.} 0$.

Theorem 5 (Regularized Regression) *Assume that the simulation error sequence $\{E_k\}$ satisfies for some $p > 2$ that*

$$\limsup_{k\to\infty} k^p\, \mathbf{E}\big[\|E_k\|^{2p}\big] < \infty.$$

The iteration

$$x_{k+1} = x_k - (A_k'\Sigma^{-1}A_k + \beta I)^{-1} A_k'\Sigma^{-1}(A_k x_k - b_k)$$

generates $\{x_k\}$ such that $Ax_k - b \xrightarrow{a.s.} 0$.

Summary

Algorithms	Residuals $r_k = Ax_k - b$	Iterates x_k and $\hat{x}_k = \Pi_k x_k$
General iteration	r_k may diverge.	x_k and $\hat{x}_k$ may diverge.
Stabilized iteration	$r_k \xrightarrow{a.s.} 0$.	x_k and $\hat{x}_k$ converge.
Nullspace-consistent iteration	$r_k \xrightarrow{a.s.} 0$.	x_k may diverge, but $\hat{x}_k$ converges.
Proximal iteration with quadratic regularization	$r_k \xrightarrow{a.s.} 0$.	x_k may diverge, but $\hat{x}_k$ converges.

Acknowledgement

Work supported by the Air Force Grant FA9550-10-1-0412 and by the LANL Information of Science and Technology Institute.

References

[WaB12a] Wang, M., and Bertsekas, D. P., 2012. "Stabilization of Stochastic Iterative Methods for Singular and Nearly Singular Linear Systems, " LIDS-P-2878, MIT, Cambridge, M. A.

[WaB12b] Wang, M., and Bertsekas, D. P., 2012. "On the Convergence of Simulation-Based Iterative Methods for Singular Linear Systems," LIDS-P-2879, MIT, Cambridge, M. A.