Preconditioning techniques based on the Birkhoff-von Neumann decomposition

Bora Uçar

CNRS & ENS Lyon, France

CSC 2016, 10–12 October, 2016, Albuquerque

From joint work with:

Michele Benzi
Emory Univ., US

Alex Pothen
Purdue Univ., US

Problem

Develop and investigate preconditioners for Krylov subspace methods for solving $\mathbf{A}x = b$, with $\mathbf{A}$ highly unstructured and indefinite.

How?

- Preprocess to have a doubly stochastic matrix (whose row and column sums are one).
- Using this doubly stochastic matrix, select some fraction of some of the nonzeros of $\mathbf{A}$ to be included in the preconditioner.

Why?

Preconditioners can be applied to vectors by a number of highly concurrent steps, where the number of steps is controlled by the user.

Main ingredients: Birkhoff-von Neumann (BvN) decomposition, and matrix splitting of the form $\mathbf{A} = \mathbf{M} - \mathbf{N}$.

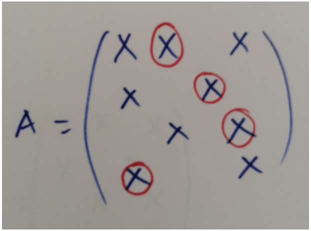
Contributions

- Sufficient conditions when such a splitting is convergent
- Specialized solvers for $\mathbf{M}y = z$ when these conditions are met.
- Use as preconditioners (e.g., with LU decomposition $\mathbf{M}$: it is of the type "complete decomposition of an incomplete matrix" as opposed to incomplete decomposition of a complete matrix).

Context

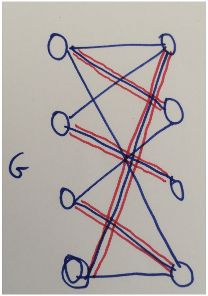
Matrix view

- Permutation matrix: An $n \times n$ matrix with exactly one 1 in each row and in each column (other entries are 0)



Bipartite graph view

- Perfect matching in $(\mathcal{R} \cup \mathcal{C}, E)$: a set of n edges no two share a common vertex.



Context

An $n \times n$ matrix $\mathbf{A}$ is doubly stochastic if $a_{ij} \geq 0$, and row sums and column sums are 1.

A doubly stochastic matrix has perfect matchings touching all of its nonzeros.

Birkhoff's Theorem: $\mathbf{A}$ is a doubly stochastic matrix

There exist $\alpha_1, \alpha_2, \ldots, \alpha_k \in (0, 1)$ with $\sum_{i=1}^{k} \alpha_i = 1$ and permutation matrices $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_k$ such that:

$$\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k \,.$$

- Also called Birkhoff-von Neumann (BvN) decomposition.
- Not unique, neither k, nor $\mathbf{P}_i$s in general.
- Finding the minimum number k of permutation matrices is NP hard.

Motivation

Consider solving $\alpha \mathbf{P} x = b$ for x where $\mathbf{P}$ is a permutation matrix.

$$\alpha \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} \text{ yields } \begin{array}{l} x_4 = b_1/\alpha \\ x_3 = b_2/\alpha \\ x_1 = b_3/\alpha \\ x_2 = b_4/\alpha \end{array}$$

We just scale the input and write at unique (permuted) positions in the output. Should be very efficient.

Next consider solving $(\alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2)x = b$ for x.

Motivation

Consider solving $(\alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2)x = b$ for x.

Matrix splitting and stationary iterations

For an invertible $\mathbf{A} = \mathbf{M} - \mathbf{N}$ with invertible $\mathbf{M}$

$$x^{(i+1)} = \mathbf{H}x^{(i)} + c, \quad \text{where} \quad \mathbf{H} = \mathbf{M}^{-1}\mathbf{N} \quad \text{and} \quad c = \mathbf{M}^{-1}b$$

where $k = 0, 1, \ldots$ and $x^{(0)}$ is arbitrary.

- Computation: At every step, multiply with $\mathbf{N}$ and solve with $\mathbf{M}$.
- Converges to the solution of $\mathbf{A}x = b$ for any $x^{(0)}$ if and only if $\rho(\mathbf{H}) < 1$ [largest magnitude of an eigenvalue is less than 1].

Motivation

Theorem

Let $\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2$ *and* $\alpha_1 \geq \alpha_2$. *Then,* $\mathbf{A}$ *is invertible if*

(i) $\alpha_1 \neq \alpha_2$,

(ii) $\alpha_1 = \alpha_2$ *and all connected components of* $G_{\mathbf{A}}$ *have an odd number of rows (and columns). If any such block is of even order,* $\mathbf{A}$ *is singular.*

Define the splitting $\mathbf{A} = \alpha_1 \mathbf{P}_1 - (-\alpha_2 \mathbf{P}_2)$.

The iterations are convergent with the rate α_2/α_1 for $\alpha_1 > \alpha_2$.

Next generalize to more than two permutation matrices.

Motivation: Let's generalize to solve $\mathbf{A}x = b$

Let $\mathbf{A} = \alpha_1\mathbf{P}_1 + \alpha_2\mathbf{P}_2 + \cdots + \alpha_k\mathbf{P}_k$ be a BvN.

Assume $\alpha_1 \geq \cdots \geq \alpha_k$. Pick an integer r between 1 and $k-1$ and split $\mathbf{A}$ as $\mathbf{A} = \mathbf{M} - \mathbf{N}$ where

$$\mathbf{M} = \alpha_1\mathbf{P}_1 + \cdots + \alpha_r\mathbf{P}_r, \quad \mathbf{N} = -\alpha_{r+1}\mathbf{P}_{r+1} - \cdots - \alpha_k\mathbf{P}_k.$$

($\mathbf{M}$ and $-\mathbf{N}$ are doubly substochastic matrices.)

Computation: At every step $\mathbf{M}^{-1}\mathbf{N}x^{(i)}$

- multiply with $\mathbf{N}$ ($k-r$ parallel steps).
- apply $\mathbf{M}^{-1}$ (or solves with the doubly stochastic matrix $\frac{1}{1-\sum_{i=r+1}^{k}\alpha_i}\mathbf{M}$); a recursive solver.

Motivation: Let's generalize more

Splitting $\mathbf{A} = \mathbf{M} - \mathbf{N}$ where

$$\mathbf{M} = \alpha_1 \mathbf{P}_1 + \cdots + \alpha_r \mathbf{P}_r, \quad \mathbf{N} = -\alpha_{r+1}\mathbf{P}_{r+1} - \cdots - \alpha_k \mathbf{P}_k.$$

Theorem

A sufficient condition for $\mathbf{M} = \sum_{i=1}^{r} \alpha_i \mathbf{P}_i$ to be invertible: α_1 is greater than the sum of the remaining ones.

Theorem

Suppose that α_1 is greater than the sum of all the other α_i. Then $\rho(\mathbf{M}^{-1}\mathbf{N}) < 1$ and the stationary iterative method converges for all x^0 to the unique solution of $\mathbf{A}x = b$.

This is a sufficient condition; ...and it is rather restrictive in practice.☹

Motivation: Let's generalize to any A

$\mathbf{M}$ as a preconditioner for a Krylov subspace method like GMRES.

... need to generalize to matrices with negative and positive entries.

Scaling fact

Any nonnegative matrix $\mathbf{A}$ with total support can be scaled with two (unique) positive diagonal matrices $\mathbf{R}$ and $\mathbf{C}$ such that $\mathbf{RAC}$ is doubly stochastic.

Let $\mathbf{A}$ be $n \times n$ with total support and positive and negative entries.

$\mathbf{B} = \text{abs}(\mathbf{A})$ is nonnegative and $\mathbf{RBC}$ is doubly stochastic.

We can write $\mathbf{RBC} = \sum \alpha_i \mathbf{P}_i$.

Motivation: Let's generalize to any A

$\mathbf{B} = \text{abs}(\mathbf{A})$ and $\mathbf{RBC} = \sum_i^k \alpha_i \mathbf{P}_i$.

$$\mathbf{RAC} = \sum_i^k \alpha_i \mathbf{Q}_i \,.$$

where $\mathbf{Q}_i = [q_{jk}^{(i)}]_{n\times n}$ is obtained from $\mathbf{P}_i = [p_{jk}^{(i)}]_{n\times n}$ as follows:

$$q_{jk}^{(i)} = \text{sgn}(a_{jk}) p_{jk}^{(i)} \,.$$

Generalizing Birkhoff–von Neumann decomposition

Any (real) matrix $\mathbf{A}$ with total support can be written as a convex combination of a set of signed, scaled permutation matrices.

We can then use the same construct to define $\mathbf{M}$ (for splitting or for defining the preconditioner).

Motivation: Let's generalize to any A (for having a special solver)

Select only a few $\alpha_i \mathbf{P}_i$ from the BvN decomposition:

$$\begin{aligned} \mathbf{A} &= \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k \,. \\ \mathbf{M} &= \alpha_1 \mathbf{P}_1 + \alpha_3 \mathbf{P}_3 + \cdots + \alpha_i \mathbf{P}_i \end{aligned}$$

We have a greedy algorithm which finds α_i in non-increasing order.

Find first 10–15 $\alpha_i \mathbf{P}_i$, take α_1 (the largest) into $\mathbf{M}$, and add the others as long as α_1 is greater than their sum.

Experiments

- All chemical, real, square matrices from the UFL collection (70 matrices) — nasty for Krylov subspace methods. Work with the largest fully indecomposable block.
- Two sets: Nonnegative and general
- (F)Gmres at most 3K iterations with 1.0e-6. Check output for accuracy ($>$ 1.0e-4 is not accurate).
- Scaling algorithm of Knight and Ruiz'13 [IMA J. Numer. Anal.], with tolerance 1.0e-8.
- ILU with all suggested preprocessing.
- LU of BvN based preconditioners with differing number of permutation matrices, and the specialized solver (select $\alpha_i\mathbf{P}_i$ in such a way that $\alpha_1 > \sum$ the rest).

Experiments

Number of failed instances

	nonnegative	general
ILU(0)	47	47
$LU(BvN)_1$	24	19
$LU(BvN)_2$	12	13
$LU(BvN)_4$	25	33
$LU(BvN)_{16}$	33	33
BvN-Solver	6	4

About 7 matrices for the solver.

Re-checked earlier results (be watchful of warnings)

- ILU fails in 17 out of 28 nonnegative matrices, and in 14 general matrices.
- BvN-solver fails in 8 nonnegative, and in 4 general matrices

Insights

- The better scaling, the better the BvN decomposition as an approximation.
- Inner splitting based solver can be used with less accuracy than the outer solver (fgmres).
- Usually, the more matrices in $\mathbf{M}$, the better the number of iterations (not for the matrices for which scaling algorithms have issues).

Experiments (running times)

One of the hard cases (for scaling even) 'bayer08', $n = 1734$, nnz=17363.

- Scaling (15K iters): 1.52 seconds
- BvN decomposition (finds 518 matchings): 0.70.
- BvN-Solve: 162 iters, 3.27 seconds (7 matchings)
- ILU: 35 iters. (set up time < 0.01 seconds), 0.13 seconds;

Conclusions

What?: Find a set of permutation matrices with scaled entries to define a preconditioner.

Why?: Exposes parallelism in applying the preconditioner.

How?: Scale the matrices and use Birkhoff–von Neumann decomposition; even for matrices with positive and negative entries.

Future work: Reduce the running time of the construction; parallel experiments.

Thank you for your attention.

References I

- M. Benzi, A. Pothen, and B. Uçar, Preconditioning techniques based on the Birkhoff–von Neumann decomposition, Tech. Rep. RR-8914, Inria Grenobl—Rhône-Alpes, 2016.
- R. A. Brualdi, The diagonal hypergraph of a matrix (bipartite graph), Discrete Mathematics 27 (2) (1979) 127–147.
- R. A. Brualdi, Notes on the Birkhoff algorithm for doubly stochastic matrices, Canadian Mathematical Bulletin, 25 (1982), pp. 191–199.
- R. A. Brualdi and P. M. Gibson, Convex polyhedra of doubly stochastic matrices: I. Applications of the permanent function, Journal of Combinatorial Theory, Series A, 22 (1977), pp. 194–230.
- C. S. Chang , W. J. Chen, and H. Y. Huang, On service guarantees for input buffered crossbar switches: A capacity decomposition approach by Birkhoff and von Neumann, IEEE IWQos'99, 79–86, London, UK.

References II

- I. S. Duff and J. Koster, On algorithms for permuting large entries to the diagonal of a sparse matrix, SIAM Journal on Matrix Analysis and Applications 22 (2001) 973–996.
- F. Dufossé and B. Uçar, Notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices, Linear Algebra and its Applications 497 (2016) 108–115.
- P. A. Knight, The Sinkhorn–Knopp algorithm: Convergence and applications, SIAM J. Matrix Anal. A. 30 (1) (2008) 261–275.
- P. A. Knight and D. Ruiz, A fast algorithm for matrix balancing, IMA Journal of Numerical Analysis 33 (3) (2013) 1029–1047.
- P. A. Knight, D. Ruiz, and B. Uçar, A symmetry preserving algorithm for matrix scaling, SIAM J. Matrix Anal. A. 35 (3) (2014) 931–955.
- M. Marcus and R. Ree, Diagonals of doubly stochastic matrices, The Quarterly Journal of Mathematics 10 (1) (1959) 296–302.

References III

- A. Pothen and C.-J. Fan, Computing the block triangular form of a sparse matrix, ACM T. Math. Software 16 (4) (1990) 303–324.
- D. Ruiz, A scaling algorithm to equilibrate both row and column norms in matrices, Tech. Rep. TR-2001-034, RAL (2001).
- R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, Pacific J. Math. 21 (1967) 343–348.
- D. de Werra, Variations on the Theorem of Birkhoff–von Neumann and extensions, Graphs and Combinatorics, 19 (2003), 263–278.
- D. de Werra, Partitioning the edge set of a bipartite graph into chain packings: Complexity of some variations, Linear Algebra and its Applications, 368 (2003) 315–327.