Technical Report 2011-1

SPIKE – A Hybrid Algorithm for Large Banded Systems

Toby Heyn, Dan Negrut

August 16, 2011

Abstract

This paper describes the SPIKE algorithm for solving large banded linear systems using a divide-and-conquer approach. The algorithm works by first partitioning the matrix into sub-matrices which are factored in parallel, then solving a reduced coupling problem recursively, a then recovering the solution to the original problem. This algorithm should be combined with re-ordering strategies to reduce the bandwidth of the matrix and to make it diagonally "heavy", or diagonally dominant if possible. Additionally, this algorithm can be used as a preconditioner if it is used inside of an outer Krylov iteration.

Contents

1. Introduction

This document describes an algorithm which can be used to solve large sparse banded linear systems in parallel. The algorithm, called SPIKE [1-3], is extremely scalable, solving linear systems in a divide-and-conquer approach. From a high vantage point, the method has several stages: reorder the matrix to obtain, if possible, a dense banded matrix that is diagonally "heavy" or, better yet, diagonally dominant [4]; partition the banded matrix into p diagonal blocks interconnected by small coupling sub-matrices; concurrently produce the p LU factorizations of the diagonal blocks and set up the reduced order coupling problem; and solve the reduced order coupling problem to recover the solution of the linear system. In various implementations this approach has been shown in preliminary numerical experiments to scale to thousands of processors, solve systems with millions of unknowns, and outperform several other sparse solvers by one order of magnitude or more [2, 3, 5-9].

2. SPIKE Algorithm

Let $\mathbf{Ax} = \mathbf{f}$, $\mathbf{A} \in \mathbb{R}^n$, be a nonsymmetric banded linear system. SPIKE relies on the factorization $\mathbf{A} = \mathbf{D} \times \mathbf{S}$, where $\mathbf{D}$ is a block-diagonal matrix and $\mathbf{S}$ is called the spike matrix (see Figure 1). The number of diagonal blocks in $\mathbf{D}$ is p (in this case, $p=4$). Note that $\mathbf{A}_i \in \mathbb{R}^{n_i \times n_i}$; $\mathbf{x}_i, \mathbf{f}_i, \mathbf{g}_i \in \mathbb{R}^{n_i}$; $\sum_{i=1}^{p} n_i = n$ are induced by partitioning.

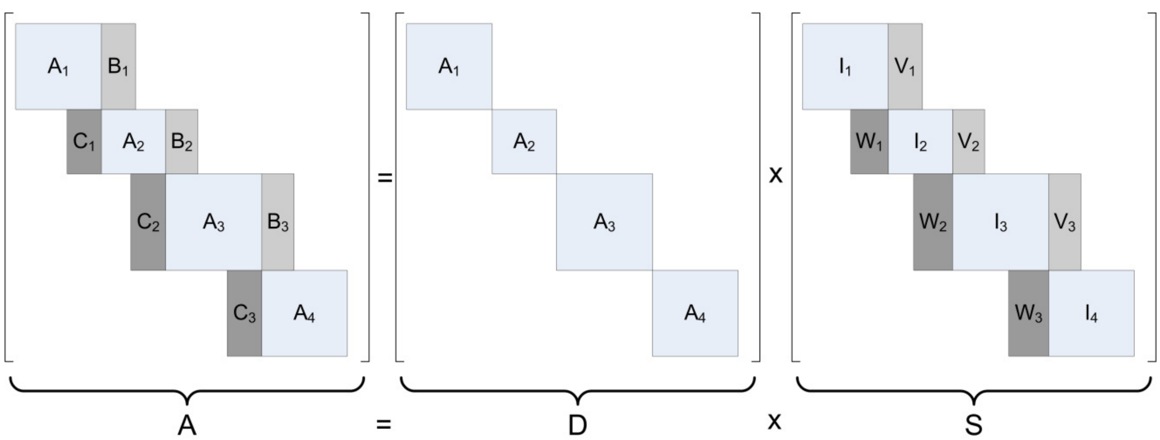

Figure 1: SPIKE factorization $\mathbf{A} = \mathbf{D} \times \mathbf{S}$.

The SPIKE factorization leads to the following set of equations.

$$\begin{aligned} \mathbf{Dg} &= \mathbf{f} \\ \mathbf{Sx} &= \mathbf{g} \end{aligned} \tag{1}$$

The basic SPIKE algorithm consists of the following four steps:

(***S1***) Concurrently obtain the LU-factorization without pivoting of the diagonal blocks $\mathbf{A}_i$; i.e., $\mathbf{A}_i = \mathbf{L}_i\mathbf{U}_i$, $i = 1,\ldots,p$

(***S2***) Assemble the spike matrix $\mathbf{S}$; i.e., concurrently compute the spikes $\mathbf{V}_i \in \mathbb{R}^{n_i \times v_i}$, $\mathbf{W}_i \in \mathbb{R}^{n_{i+1} \times w_i}$, $i = 1,\ldots,p-1$ and the right hand side $\mathbf{g}_i \in \mathbb{R}^{n_i}$, $i = 1,\ldots,p$. To this end, using the LU factorizations of $\mathbf{A}_i$, solve

$$\begin{aligned}
&\mathbf{L}_1\mathbf{U}_1\left[\mathbf{V}_1,\mathbf{g}_1\right] = \left[\mathbf{B}_1,\mathbf{f}_1\right] \\
&\mathbf{L}_i\mathbf{U}_i\left[\mathbf{V}_i,\mathbf{W}_{i-1},\mathbf{g}_i\right] = \left[\mathbf{B}_i,\mathbf{C}_{p-1},\mathbf{f}_i\right], \quad i = 2,\ldots,p-1 \\
&\mathbf{L}_p\mathbf{U}_p\left[\mathbf{V}_{p-1},\mathbf{g}_p\right] = \left[\mathbf{C}_{p-1},\mathbf{f}_p\right]
\end{aligned} \tag{2}$$

(***S3***) Solve the reduced block-tridiagonal system,

$$\begin{bmatrix}
\mathbf{R}_1 & \mathbf{M}_1 & & & & & \\
\mathbf{N}_1 & \mathbf{R}_2 & \mathbf{M}_2 & & & & \\
 & \mathbf{N}_2 & \ddots & \ddots & & & \\
 & & \ddots & \mathbf{R}_{p-3} & \mathbf{M}_{p-3} & & \\
 & & & \mathbf{N}_{p-3} & \mathbf{R}_{p-2} & \mathbf{M}_{p-2} \\
 & & & & \mathbf{N}_{p-2} & \mathbf{R}_{p-1}
\end{bmatrix} \times \begin{bmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \\ \vdots \\ \hat{\mathbf{x}}_{p-3} \\ \hat{\mathbf{x}}_{p-2} \\ \hat{\mathbf{x}}_{p-1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \vdots \\ \hat{\mathbf{g}}_{p-3} \\ \hat{\mathbf{g}}_{p-2} \\ \hat{\mathbf{g}}_{p-1} \end{bmatrix} \tag{3}$$

where $\mathbf{R}_i = \begin{bmatrix} \mathbf{I} & \mathbf{V}_i^a \\ \mathbf{W}_i^b & \mathbf{I} \end{bmatrix}$, $\mathbf{M}_i = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{i+1}^b \end{bmatrix}$, $\mathbf{N}_i = \begin{bmatrix} \mathbf{W}_i^a & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $\hat{\mathbf{x}}_i = \begin{bmatrix} \mathbf{x}_i^a \\ \mathbf{x}_i^b \end{bmatrix}$, $\hat{\mathbf{g}}_i = \begin{bmatrix} \mathbf{g}_i^a \\ \mathbf{g}_i^b \end{bmatrix}$

and $\mathbf{V}_i^a \in \mathbb{R}^{w_i \times v_i}$, $\mathbf{V}_i^b \in \mathbb{R}^{v_{i-1} \times v_i}$, $\mathbf{W}_i^a \in \mathbb{R}^{w_{i+1} \times w_i}$, $\mathbf{W}_i^b \in \mathbb{R}^{v_i \times w_i}$ are respectively the bottom w_i rows of $\mathbf{V}_i$, the top v_{i-1} rows of $\mathbf{V}_i$, the bottom w_{i+1} rows of $\mathbf{W}_i$, and the top v_i rows of $\mathbf{W}_i$.

(***S4***) Recover the solution of the original linear system by concurrently computing

$$\begin{aligned}
&\mathbf{x}_1' = \mathbf{g}_1' - \mathbf{V}_1'\mathbf{x}_1^b \\
&\mathbf{x}_i' = \mathbf{g}_k' - \mathbf{W}_{i-1}'\mathbf{x}_{i-1}^a - \mathbf{V}_i'\mathbf{x}_i^b, \quad i = 2,\ldots,p-1 \\
&\mathbf{x}_p' = g_p' - \mathbf{W}_{p-1}'\mathbf{x}_{p-1}^a
\end{aligned} \tag{4}$$

Then the solution of the original linear system is assembled as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{bmatrix} \text{ where } \mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_1^a \end{bmatrix}, \mathbf{x}_p = \begin{bmatrix} \mathbf{x}_{p-1}^b \\ \mathbf{x}_p' \end{bmatrix}, \text{ and } \mathbf{x}_i = \begin{bmatrix} \mathbf{x}_{i-1}^b \\ \mathbf{x}_i' \\ \mathbf{x}_i^a \end{bmatrix}, i = 2,\ldots,p-1.$$

2.1. Recursive solution of reduced system

The system in Equation 3 can be solved directly, or via a recursive method which eliminates pairs of equations associated with partition boundaries. Non-adjacent partition boundaries can be eliminated simultaneously, meaning that the *p-1* boundaries of a system with *p* partitions can be eliminated in $\log_2\left(p\right)$ steps.

2.1.1. Elimination of a single border

Border $\left(k\right)$ is eliminated as follows. The relevant equations from Equation 3 (associated with borders $\left(k-1\right)$, $\left(k\right)$, and $\left(k+1\right)$) are as follows:

$$\begin{aligned}
(k\text{-}1)_b:&\ \mathbf{W}_{k-1}^{b}\mathbf{x}_{k-1}^{a}+\mathbf{x}_{k-1}^{b}+\mathbf{V}_{k}^{b}\mathbf{x}_{k}^{b}=\mathbf{g}_{k-1}^{b}\\
(k)_{\mathrm{a}}:&\ \mathbf{W}_{k-1}^{a}\mathbf{x}_{k-1}^{a}+\mathbf{x}_{k}^{a}+\mathbf{V}_{k}^{a}\mathbf{x}_{k}^{b}=\mathbf{g}_{k}^{a}\\
(k)_b:&\ \mathbf{W}_{k}^{b}\mathbf{x}_{k}^{a}+\mathbf{x}_{k}^{b}+\mathbf{V}_{k+1}^{b}\mathbf{x}_{k+1}^{b}=\mathbf{g}_{k}^{b}\\
(k+1)_{\mathrm{a}}:&\ \mathbf{W}_{k}^{a}\mathbf{x}_{k}^{a}+\mathbf{x}_{k+1}^{a}+\mathbf{V}_{k+1}^{a}\mathbf{x}_{k+1}^{b}=\mathbf{g}_{k+1}^{a}
\end{aligned}$$

Equations $\left(k\right)_a$ and $\left(k\right)_b$ above can be solved simultaneously for $\mathbf{x}_k^a$ and $\mathbf{x}_k^b$.

$$\mathbf{x}_{k}^{a}=\left(\mathbf{D}_{k}^{ab}\right)^{-1}\left(\mathbf{g}_{k}^{a}-\mathbf{V}_{k}^{a}\mathbf{g}_{k}^{b}\right)-\left(\mathbf{D}_{k}^{ab}\right)^{-1}\mathbf{W}_{k-1}^{a}\mathbf{x}_{k-1}^{a}+\left(\mathbf{D}_{k}^{ab}\right)^{-1}\mathbf{V}_{k}^{a}\mathbf{V}_{k+1}^{b}\mathbf{x}_{k+1}^{b}$$

$$\text{where } \mathbf{D}_{k}^{ab}=\mathbf{I}-\mathbf{V}_{k}^{a}\mathbf{W}_{k}^{b}$$

$$\mathbf{x}_{k}^{b}=\left(\mathbf{D}_{k}^{ba}\right)^{-1}\left(\mathbf{g}_{k}^{b}-\mathbf{W}_{k}^{b}\mathbf{g}_{k}^{a}\right)-\left(\mathbf{D}_{k}^{ba}\right)^{-1}\mathbf{V}_{k+1}^{b}\mathbf{x}_{k+1}^{b}+\left(\mathbf{D}_{k}^{ba}\right)^{-1}\mathbf{W}_{k}^{b}\mathbf{W}_{k-1}^{a}\mathbf{x}_{k-1}^{a}$$

$$\text{where } \mathbf{D}_{k}^{ba}=\mathbf{I}-\mathbf{W}_{k}^{b}\mathbf{V}_{k}^{a}$$

Now, substituting $\mathbf{x}_k^a$ into equation $\left(k+1\right)_a$ and $\mathbf{x}_k^b$ into equation $\left(k-1\right)_b$ leads to the following:

$$(k\text{-}1)_{\mathrm{b}}:\ \left[\mathbf{W}_{k-1}^{b}+\mathbf{V}_{k}^{b}\left(\mathbf{D}_{k}^{ba}\right)^{-1}\mathbf{W}_{k}^{b}\mathbf{W}_{k-1}^{a}\right]\mathbf{x}_{k-1}^{a}+\mathbf{x}_{k-1}^{b}+\mathbf{V}_{k}^{b}\left(-\mathbf{D}_{k}^{ba}\right)^{-1}\mathbf{V}_{k+1}^{b}\mathbf{x}_{k+1}^{b}=\mathbf{g}_{k-1}^{b}-\mathbf{V}_{k}^{b}\left(\mathbf{D}_{k}^{ba}\right)^{-1}\left(\mathbf{g}_{k}^{b}-\mathbf{W}_{k}^{b}\mathbf{g}_{k}^{a}\right)$$

$$(k+1)_{a}:\ \mathbf{W}_{k}^{a}\left(-\mathbf{D}_{k}^{ab}\right)^{-1}\mathbf{W}_{k-1}^{a}\mathbf{x}_{k-1}^{a}+\mathbf{x}_{k+1}^{a}+\left[\mathbf{V}_{k+1}^{a}+\mathbf{W}_{k}^{a}\left(\mathbf{D}_{k}^{ab}\right)^{-1}\mathbf{V}_{k}^{a}\mathbf{V}_{k+1}^{b}\right]\mathbf{x}_{k+1}^{b}=\mathbf{g}_{k+1}^{a}-\mathbf{W}_{k}^{a}\left(\mathbf{D}_{k}^{ab}\right)^{-1}\left(\mathbf{g}_{k}^{a}-\mathbf{V}_{k}^{a}\mathbf{g}_{k}^{b}\right)$$

This process eliminated the unknowns (and the equations) associated with border $\left(k\right)$. It also preserved the nature of the problem; i.e., the structure of the reduced system remains the same. This becomes obvious if the following notation is used:

$$\mathbf{W}_{k-1}^{a}\boxed{k} = \mathbf{W}_{k}^{a}\left(-\mathbf{D}_{k}^{ab}\right)^{-1}\mathbf{W}_{k-1}^{a}$$
$$\mathbf{W}_{k-1}^{b}\boxed{k} = \mathbf{W}_{k-1}^{b} + \mathbf{V}_{k}^{b}\left(\mathbf{D}_{k}^{ba}\right)^{-1}\mathbf{W}_{k}^{b}\mathbf{W}_{k-1}^{a}$$
$$\mathbf{V}_{k+1}^{a}\boxed{k} = \mathbf{V}_{k+1}^{a} + \mathbf{W}_{k}^{a}\left(\mathbf{D}_{k}^{ab}\right)^{-1}\mathbf{V}_{k}^{a}\mathbf{V}_{k+1}^{b}$$
$$\mathbf{V}_{k+1}^{b}\boxed{k} = \mathbf{V}_{k}^{b}\left(-\mathbf{D}_{k}^{ba}\right)^{-1}\mathbf{V}_{k+1}^{b}$$
$$\mathbf{g}_{k+1}^{a}\boxed{k} = \mathbf{g}_{k+1}^{a} - \mathbf{W}_{k}^{a}\left(\mathbf{D}_{k}^{ab}\right)^{-1}\left(\mathbf{g}_{k}^{a} - \mathbf{V}_{k}^{a}\mathbf{g}_{k}^{b}\right)$$
$$\mathbf{g}_{k-1}^{b}\boxed{k} = \mathbf{g}_{k-1}^{b} - \mathbf{V}_{k}^{b}\left(\mathbf{D}_{k}^{ba}\right)^{-1}\left(\mathbf{g}_{k}^{b} - \mathbf{W}_{k}^{b}\mathbf{g}_{k}^{a}\right)$$

Then, the relevant part of the reduced system looks like

$$(k\text{-}1)_{\mathrm{b}}:\ \mathbf{W}_{k-1}^{b}\boxed{k}\mathbf{x}_{k-1}^{a} + \mathbf{x}_{k-1}^{b} + \mathbf{V}_{k+1}^{b}\boxed{k}\mathbf{x}_{k+1}^{b} = \mathbf{g}_{k-1}^{b}\boxed{k}$$
$$(k\text{+}1)_{\mathrm{a}}:\ \mathbf{W}_{k-1}^{a}\boxed{k}\mathbf{x}_{k-1}^{a} + \mathbf{x}_{k+1}^{a} + \mathbf{V}_{k+1}^{a}\boxed{k}\mathbf{x}_{k+1}^{b} = \mathbf{g}_{k+1}^{a}\boxed{k}$$

Eliminating border $\left(k\right)$ impacts only equations $\left(k+1\right)_a$ and $\left(k-1\right)_b$. Therefore, any two non-adjacent borders can be eliminated simultaneously.

2.1.2. Generalization and recursion

The purpose of this section is to describe the notation and formulation after *S* border elimination stages. Assume that at this point we want to remove border (k). Let the adjacent borders be border (e) above and border (p) below. The following notation is used to indicate, for example, that the border above border (k) after *S* elimination stages is border (e).

$$e = A\left(k \mid S\right)$$
$$p = B\left(k \mid S\right)$$

Note that using this notation, the following properties hold true:

$$A\left(k \mid 0\right) = k-1,\ B\left(k \mid 0\right) = k+1$$
$$A\left(B\left(k \mid S\right) \mid S+1\right) = A\left(k \mid S\right),\ B\left(A\left(k \mid S\right) \mid S+1\right) = B\left(k \mid S\right)$$
$$A\left(k \mid S\right) < B\left(k \mid S\right)$$

Now, remove border (k) in elimination stage *S+1*. This results in the following new terms.

$$\mathbf{W}_e^{b|S+1} = \mathbf{W}_e^{b|S} + \mathbf{V}_k^{b|S}\left(\mathbf{D}_k^{ba|S}\right)^{-1}\mathbf{W}_k^{b|S}\mathbf{W}_e^{a|S} \qquad \mathbf{W}_e^{a|S+1} = \mathbf{W}_k^{a|S}\left(-\mathbf{D}_k^{ab|S}\right)^{-1}\mathbf{W}_e^{a|S}$$

$$\mathbf{V}_p^{b|S+1} = \mathbf{V}_k^{b|S}\left(-\mathbf{D}_k^{ba|S}\right)^{-1}\mathbf{V}_p^{b|S} \qquad \mathbf{V}_p^{a|S+1} = \mathbf{V}_p^{a|S} + \mathbf{W}_k^{a|S}\left(\mathbf{D}_k^{ab|S}\right)^{-1}\mathbf{V}_k^{a|S}\mathbf{V}_p^{b|S}$$

$$\mathbf{g}_e^{b|S+1} = \mathbf{g}_e^{b|S} - \mathbf{V}_k^{b|S}\left(\mathbf{D}_k^{ba|S}\right)^{-1}\left(\mathbf{g}_k^{b|S} - \mathbf{W}_k^{b|S}\mathbf{g}_k^{a|S}\right) \quad \mathbf{g}_p^{a|S+1} = \mathbf{g}_p^{a|S} - \mathbf{W}_k^{a|S}\left(\mathbf{D}_k^{ab|S}\right)^{-1}\left(\mathbf{g}_k^{a|S} - \mathbf{V}_k^{a|S}\mathbf{g}_k^{b|S}\right)$$

$$\text{where } \mathbf{D}_k^{ab|S} = \mathbf{I} - \mathbf{V}_k^{a|S}\mathbf{W}_k^{b|S} \text{ and } \mathbf{D}_k^{ba|S} = \mathbf{I} - \mathbf{W}_k^{b|S}\mathbf{V}_k^{a|S}$$

Now, equation $(e)_b$ and equation $(p)_a$ can be written as

$$(e)_{\mathrm{b}}:\ \mathbf{W}_e^{b|S+1}\mathbf{x}_e^a + \mathbf{x}_e^b + \mathbf{V}_p^{b|S+1}\mathbf{x}_p^b = \mathbf{g}_p^{b|S+1}$$

$$(p)_{\mathrm{a}}:\ \mathbf{W}_e^{a|S+1}\mathbf{x}_e^a + \mathbf{x}_p^a + \mathbf{V}_p^{a|S+1}\mathbf{x}_p^b = \mathbf{g}_p^{a|S+1}$$

Note that any non-consecutive boundaries can be eliminated simultaneously. Best practice is to eliminate alternating boundaries during elimination stage *S*. The following notation will be useful.

- Let $\mathcal{V}^S = \left\{i : \mathbf{x}_i^a, \mathbf{x}_i^b \text{ are variables in the reduced system after stage } S\right\}$,
- $\mathcal{E}^S = \left\{i : \mathbf{x}_i^a, \mathbf{x}_i^b \text{ are eliminated from the reduced system during stage } S+1\right\}$.
- Note that $\left|j-k\right| > 1\ \forall\ j,k \in \mathcal{E}^S$
- Note that $\mathcal{V}^0 = \left\{1,\ldots,p-1\right\}$
- Note that $\mathcal{V}^{S+1} = \mathcal{V}^S - \mathcal{E}^S$

2.1.3. Details

This section describes some details and special cases of the recursive algorithm.

I. Assume that at stage S there is no border above border (k); i.e., $A\left(k \mid S\right) = \varnothing$, and let $B\left(k \mid S\right) = p$. Then, because $\mathbf{V}_k^{b|S} \equiv \mathbf{0}$ we can see that $\mathbf{V}_p^{b|S+1} \equiv \mathbf{0}$.

II. Assume that at stage S there is no border below border (k); i.e., $B\left(k \mid S\right) = \varnothing$, and let $A\left(k \mid S\right) = e$. Then, because $\mathbf{W}_k^{a|S} \equiv \mathbf{0}$ we can see that $\mathbf{W}_e^{a|S+1} \equiv \mathbf{0}$.

III. Note that we need to compute factorizations for both $\mathbf{D}_k^{ab|S}$ and $\mathbf{D}_k^{ba|S}$. Instead, compute only the factorization of the smaller matrix and use the Woodbury matrix identity, $\left(\mathbf{I} - \mathbf{PQ}\right)^{-1} = \mathbf{I} + \mathbf{P}\left(\mathbf{I} - \mathbf{QP}\right)^{-1}\mathbf{Q}$. To this end, if $\mathbf{V}_k^{a|S}$ has a smaller or equal number of rows than $\mathbf{W}_k^{b|S}$ then let $\mathbf{Q} = \mathbf{V}_k^{a|S}$ and $\mathbf{P} = \mathbf{W}_k^{b|S}$. If $\mathbf{V}_k^{a|S}$ has more rows than $\mathbf{W}_k^{b|S}$ then let $\mathbf{Q} = \mathbf{W}_k^{b|S}$ and $\mathbf{P} = \mathbf{V}_k^{a|S}$. Then, factorize $\mathbf{I} - \mathbf{QP}$ so that terms like

$\left(\mathbf{I}-\mathbf{QP}\right)^{-1}\mathbf{X}$ are easy to compute. Then, terms like $\bar{\mathbf{Y}}=\left(\mathbf{I}-\mathbf{PQ}\right)^{-1}\mathbf{Y}$ can be computed as follows.

Compute $\mathbf{X}\equiv\mathbf{QY}$

Compute $\mathbf{Z}\equiv\left(\mathbf{I}-\mathbf{QP}\right)^{-1}\mathbf{X}$

Compute $\bar{\mathbf{Y}}=\mathbf{Y}+\mathbf{PZ}$

2.1.4. Recovery of solution

The elimination process continues until there is only one border left. Let this final border be border (h), and assume that it took *R* elimination stages to reach this point.

$$\mathcal{V}^R=\left\{h\right\},\ \mathcal{E}^R=\varnothing$$

The corresponding equations are:

$$\begin{aligned}\mathbf{x}_h^a+\mathbf{V}_h^{a|R}\mathbf{x}_h^b&=\mathbf{g}_h^{a|R}\\ \mathbf{W}_h^{b|R}\mathbf{x}_h^a+\mathbf{x}_h^b&=\mathbf{g}_h^{b|R}\end{aligned}$$

This set of equations can be solved for $\mathbf{x}_h^a$ and $\mathbf{x}_h^b$.

$$\begin{aligned}\mathbf{x}_h^a&=\left(\mathbf{D}_h^{ab|R}\right)^{-1}\left(\mathbf{g}_h^{a|R}-\mathbf{V}_h^{a|R}\mathbf{g}_h^{b|R}\right)\\ \mathbf{x}_h^b&=\left(\mathbf{D}_h^{ba|R}\right)^{-1}\left(\mathbf{g}_h^{b|R}-\mathbf{W}_h^{b|R}\mathbf{g}_h^{a|R}\right)\end{aligned}$$

Now, the recovery phase starts. Revert one stage, where $A(h\mid R-1)=c$ and $B(h\mid R-1)=d$. Now, $\mathbf{x}_c^a,\mathbf{x}_c^b$ and $\mathbf{x}_d^a,\mathbf{x}_d^b$ can be computed as follows:

$$\begin{aligned}\mathbf{x}_c^a&=\left(\mathbf{D}_c^{ab|R-1}\right)^{-1}\left(\mathbf{g}_c^{a|R-1}-\mathbf{V}_c^{a|R-1}\mathbf{g}_c^{b|R-1}\right)+\left(\mathbf{D}_c^{ab|R-1}\right)^{-1}\mathbf{V}_c^{a|R-1}\mathbf{V}_h^{b|R-1}\mathbf{x}_h^b\\ &\quad\text{where } \mathbf{D}_c^{ab|R-1}=\mathbf{I}-\mathbf{V}_c^{a|R-1}\mathbf{W}_c^{b|R-1}\\ \mathbf{x}_c^b&=\left(\mathbf{D}_c^{ba|R-1}\right)^{-1}\left(\mathbf{g}_c^{b|R-1}-\mathbf{W}_c^{b|R-1}\mathbf{g}_c^{a|R-1}\right)-\left(\mathbf{D}_c^{ba|R-1}\right)^{-1}\mathbf{V}_h^{b|R-1}\mathbf{x}_h^b\\ &\quad\text{where } \mathbf{D}_k^{ba|R-1}=\mathbf{I}-\mathbf{W}_k^{b|R-1}\mathbf{V}_k^{a|R-1}\\ \mathbf{x}_d^a&=\left(\mathbf{D}_d^{ab|R-1}\right)^{-1}\left(\mathbf{g}_d^{a|R-1}-\mathbf{V}_d^{a|R-1}\mathbf{g}_d^{b|R-1}\right)-\left(\mathbf{D}_d^{ab|R-1}\right)^{-1}\mathbf{W}_h^{a|R-1}\mathbf{x}_h^a\\ &\quad\text{where } \mathbf{D}_d^{ab|R-1}=\mathbf{I}-\mathbf{V}_d^{a|R-1}\mathbf{W}_d^{b|R-1}\\ \mathbf{x}_d^b&=\left(\mathbf{D}_d^{ba|R-1}\right)^{-1}\left(\mathbf{g}_d^{b|R-1}-\mathbf{W}_d^{b|R-1}\mathbf{g}_d^{a|R-1}\right)+\left(\mathbf{D}_d^{ba|R-1}\right)^{-1}\mathbf{W}_d^{b|R-1}\mathbf{W}_h^{a|R-1}\mathbf{x}_h^a\\ &\quad\text{where } \mathbf{D}_d^{ba|R-1}=\mathbf{I}-\mathbf{W}_d^{b|R-1}\mathbf{V}_d^{a|R-1}\end{aligned}$$

Consider the general case, when reverting from stage *S+1* to stage *S*. Note that at this point, $\mathbf{x}_i^a,\mathbf{x}_i^b$ are known $\forall\ i\in\mathcal{V}^{S+1}$, and we wish to recover $\mathbf{x}_j^a,\mathbf{x}_j^b\ \forall\ j\in\mathcal{E}^S$. This is possible because $A\left(j \mid S\right), B\left(j \mid S\right) \in \mathcal{V}^{S+1} \; \forall j \in \mathcal{E}^{S}$. Then the desired quantities can be computed $\forall \; j \in \mathcal{E}^{S}$:

$$\mathbf{x}_j^a = \left(\mathbf{D}_j^{ab|S}\right)^{-1}\left(\mathbf{g}_j^{a|S} - \mathbf{V}_j^{a|S}\mathbf{g}_j^{b|S}\right) - \left(\mathbf{D}_j^{ab|S}\right)^{-1}\mathbf{W}_{A(j|S)}^{a|S}\mathbf{x}_{A(j|S)}^{a} + \left(\mathbf{D}_j^{ab|S}\right)^{-1}\mathbf{V}_j^{a|S}\mathbf{V}_{B(j|S)}^{b|S}\mathbf{x}_{B(j|S)}^{b}$$

$$\text{where } \mathbf{D}_j^{ab|S} = \mathbf{I} - \mathbf{V}_j^{a|S}\mathbf{W}_j^{b|S}$$

$$\mathbf{x}_j^b = \left(\mathbf{D}_j^{ba|S}\right)^{-1}\left(\mathbf{g}_j^{b|S} - \mathbf{W}_j^{b|S}\mathbf{g}_j^{a|S}\right) - \left(\mathbf{D}_j^{ba|S}\right)^{-1}\mathbf{V}_{B(j|S)}^{b|S}\mathbf{x}_{B(j|S)}^{b} + \left(\mathbf{D}_j^{ba|S}\right)^{-1}\mathbf{W}_j^{b|S}\mathbf{W}_{A(j|S)}^{a|S}\mathbf{x}_{A(j|S)}^{a}$$

$$\text{where } \mathbf{D}_j^{ba|S} = \mathbf{I} - \mathbf{W}_j^{b|S}\mathbf{V}_j^{a|S}$$

3. Example

This section shows the SPIKE algorithm applied to a problem with 5 partitions (*p=5*) and 4 borders. The reduced system equations can be seen schematically in Figure 2.

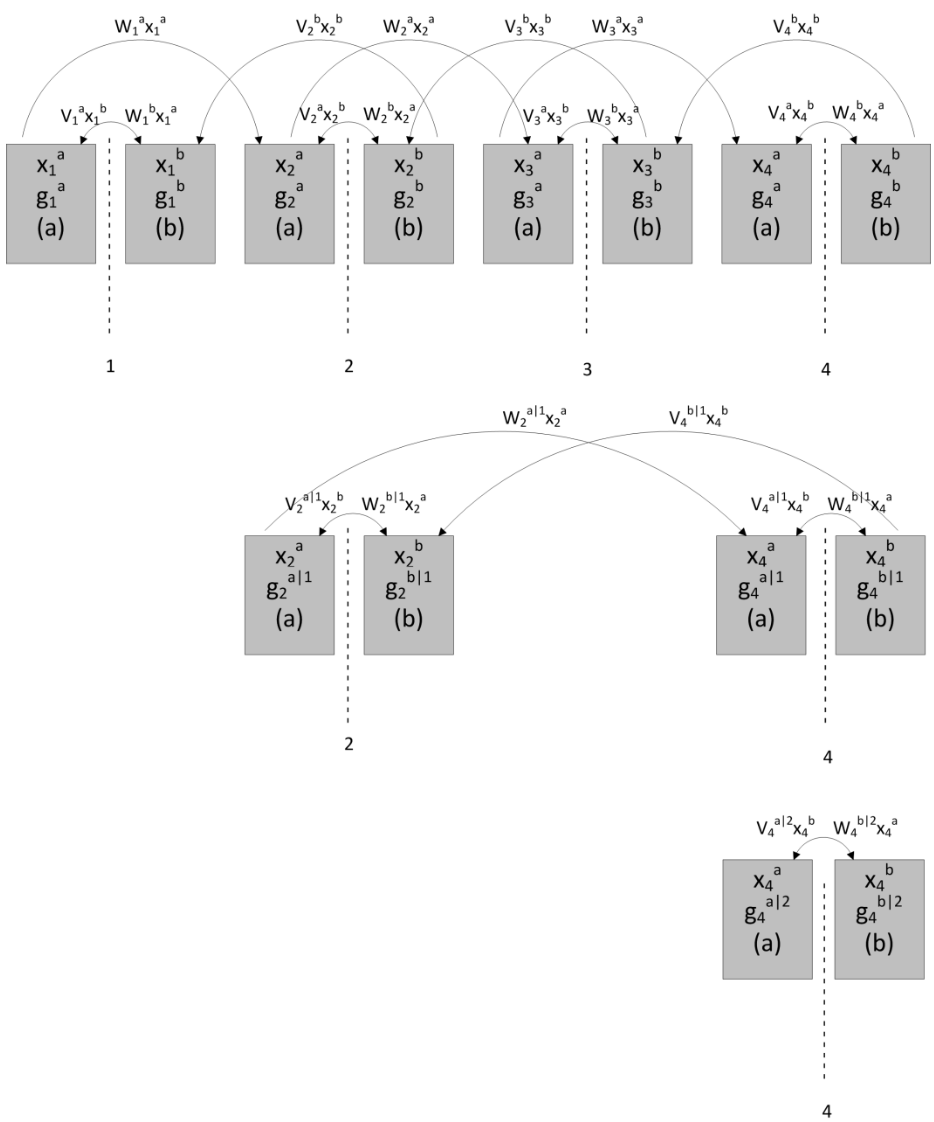

Figure 2: Schematic of reduced system equations for *p=5*.

The computation proceeds as follows:

Elimination 1:

- The first row of Figure 2 shows the full reduced system, $\mathcal{V}^0 = \{1,2,3,4\}$, and we will eliminate alternating borders starting with border (*1*), $\mathcal{E}^0 = \{1,3\}$.
- Compute $\mathbf{V}_2^{a|1}, \mathbf{V}_2^{b|1}, \mathbf{W}_2^{a|1}, \mathbf{W}_2^{b|1}$ and $\mathbf{V}_4^{a|1}, \mathbf{V}_4^{b|1}, \mathbf{W}_4^{a|1}, \mathbf{W}_4^{b|1}$

Elimination 2:

- The second row of Figure 2 shows the modified reduced system, with $\mathcal{V}^1 = \{2,4\}$, and we again eliminate alternating borders, $\mathcal{E}^1 = \{2\}$.
- Compute $\mathbf{V}_4^{a|2}, \mathbf{V}_4^{b|2}, \mathbf{W}_4^{a|2}, \mathbf{W}_4^{b|2}$

Solve single border problem:

- Now, the system has a single border left, $\mathcal{V}^2 = \{4\}, \mathcal{E}^2 = \varnothing$
- $$\mathbf{x}_4^a = \left(\mathbf{D}_4^{ab|2}\right)^{-1}\left(\mathbf{g}_4^{a|2} - \mathbf{V}_4^{a|2}\mathbf{g}_4^{b|2}\right)$$
  $$\mathbf{x}_4^b = \left(\mathbf{D}_4^{ba|2}\right)^{-1}\left(\mathbf{g}_4^{b|2} - \mathbf{W}_4^{b|2}\mathbf{g}_4^{a|2}\right)$$

Recovery 2:

- Now, we can recover $\mathbf{x}_j^a, \mathbf{x}_j^b \;\forall\; j \in \mathcal{E}^1$; i.e., we can compute $\mathbf{x}_2^a, \mathbf{x}_2^b$.
- $$\mathbf{x}_2^a = \left(\mathbf{D}_2^{ab|1}\right)^{-1}\left(\mathbf{g}_2^{a|1} - \mathbf{V}_2^{a|1}\mathbf{g}_2^{b|1}\right) + \left(\mathbf{D}_2^{ab|1}\right)^{-1}\mathbf{V}_2^{a|1}\mathbf{V}_4^{b|1}\mathbf{x}_4^b$$
  $$\mathbf{x}_2^b = \left(\mathbf{D}_2^{ba|1}\right)^{-1}\left(\mathbf{g}_2^{b|1} - \mathbf{W}_2^{b|1}\mathbf{g}_2^{a|1}\right) - \left(\mathbf{D}_2^{ba|1}\right)^{-1}\mathbf{V}_4^{b|1}\mathbf{x}_4^b$$

Recovery 1:

- Now, we can recover $\mathbf{x}_j^a, \mathbf{x}_j^b \;\forall\; j \in \mathcal{E}^0$; i.e., we can compute $\mathbf{x}_1^a, \mathbf{x}_1^b, \mathbf{x}_3^a, \mathbf{x}_3^b$.
- $$\mathbf{x}_1^a = \left(\mathbf{D}_1^{ab}\right)^{-1}\left(\mathbf{g}_1^a - \mathbf{V}_1^a\mathbf{g}_1^b\right) + \left(\mathbf{D}_1^{ab}\right)^{-1}\mathbf{V}_1^a\mathbf{V}_2^b\mathbf{x}_2^b$$
  $$\mathbf{x}_1^b = \left(\mathbf{D}_1^{ba}\right)^{-1}\left(\mathbf{g}_1^b - \mathbf{W}_1^b\mathbf{g}_1^a\right) - \left(\mathbf{D}_1^{ba}\right)^{-1}\mathbf{V}_2^b\mathbf{x}_2^b$$
  $$\mathbf{x}_3^a = \left(\mathbf{D}_3^{ab}\right)^{-1}\left(\mathbf{g}_3^a - \mathbf{V}_3^a\mathbf{g}_3^b\right) - \left(\mathbf{D}_3^{ab}\right)^{-1}\mathbf{W}_2^a\mathbf{x}_2^a + \left(\mathbf{D}_3^{ab}\right)^{-1}\mathbf{V}_3^a\mathbf{V}_4^b\mathbf{x}_4^b$$
  $$\mathbf{x}_3^b = \left(\mathbf{D}_3^{ba}\right)^{-1}\left(\mathbf{g}_3^b - \mathbf{W}_3^b\mathbf{g}_3^a\right) - \left(\mathbf{D}_3^{ba}\right)^{-1}\mathbf{V}_4^b\mathbf{x}_4^b + \left(\mathbf{D}_3^{ba}\right)^{-1}\mathbf{W}_3^b\mathbf{W}_2^a\mathbf{x}_2^a$$

Overall Solution:

- Now, we can assemble the overall solution $\mathbf{x} = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_5^T\right]^T$
- Recover the terms $\mathbf{x}'_1, \ldots, \mathbf{x}'_5$

$$
\begin{aligned}
\mathbf{x}_1{}' &= \mathbf{g}_1{}' - \mathbf{V}_1{}'\mathbf{x}_1^b \\
\mathbf{x}_i{}' &= \mathbf{g}_i{}' - \mathbf{W}'_{i-1}\mathbf{x}_{i-1}^a - \mathbf{V}_i'\mathbf{x}_i^b, \quad i = 2,\ldots,5 \\
\mathbf{x}_5{}' &= g_5{}' - \mathbf{W}_4'\mathbf{x}_4^a
\end{aligned}
$$

- Then the solution of the original linear system is assembled as

$$
\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_5 \end{bmatrix} \text{ where } \mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_1^a \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} \mathbf{x}_4^b \\ \mathbf{x}_5' \end{bmatrix}, \text{and } \mathbf{x}_i = \begin{bmatrix} \mathbf{x}_{i-1}^b \\ \mathbf{x}_i' \\ \mathbf{x}_i^a \end{bmatrix}, i = 2,\ldots,4\,.
$$

4. Conclusions

This document describes the SPIKE algorithm which can be used to solve large sparse banded linear systems in parallel. The problem is partitioned into *p* partitions, which can be factored in parallel by *p* processes. The reduced problem, based on the *p-1* boundaries between partitions, is solved with a recursive method before the solution can be recovered.

Several aspects of this algorithm will be pursued in on-going and future work. The algorithm will be implemented with CUDA support to accelerate some steps of the algorithm such as LU-factorization of the diagonal blocks, for example. Next, re-ordering strategies will be investigated to re-organize general sparse matrices into the banded structure necessary for application of this method. These re-ordering methods should also take advantage of parallelism where possible. Finally, this algorithm will be integrated with an outer Krylov-type iteration. When SPIKE is used inside of an iterative method such as GMRES, it can act as an effective preconditioner to solve problems which are not diagonally dominant in few iterations.

References

1. Polizzi, E. and A.H. Sameh, *A parallel hybrid banded system solver: the SPIKE algorithm.* Parallel computing, 2006. **32**(2): p. 177-194.
2. Polizzi, E. and A. Sameh, *SPIKE: A parallel environment for solving banded linear systems.* Computers & Fluids, 2007. **36**(1): p. 113-120.
3. Manguoglu, M., A. Sameh, and O. Schenk, *PSPIKE: A parallel hybrid sparse linear system solver.* Euro-Par 2009 Parallel Processing, 2009: p. 797-808.
4. Golub, G. and C. van Van Loan, *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition).* 1996: The Johns Hopkins University Press.
5. Chen, S., D.J. Kuck, and A.H. Sameh, *Practical parallel band triangular system solvers.* ACM Transactions on Mathematical Software (TOMS), 1978. **4**(3): p. 270-277.
6. Lawrie, D.H. and A.H. Sameh, *The computation and communication complexity of a parallel banded system solver.* ACM Transactions on Mathematical Software (TOMS), 1984. **10**(2): p. 185-195.
7. Dongarra, J.J. and A.H. Sameh, *On some parallel banded system solvers.* Parallel Computing, 1984. **1**(3-4): p. 223-235.
8. Sameh, A.H. and V. Sarin, *Hybrid parallel linear system solvers.* International Journal of Computational Fluid Dynamics, 1999. **12**(3): p. 213-223.
9. Mikkelsen, C.C.K. and M. Manguoglu, *Analysis of the truncated SPIKE algorithm.* SIAM Journal on Matrix Analysis and Applications, 2008. **30**(4): p. 1500-1519.