A Bound on Matrix-Vector Products for $(0,1)$-Matrices via Gray Codes

Andrew A. Anda

aanda@stcloudstate.edu

St. Cloud State University, CSCI Dept.

$(0, 1)$-Matrix

A $(0, 1)$-matrix (also zero-one or Boolean) is a rectangular matrix for which each element of the matrix has the value of either 0 or 1.

$(0,1)$-Matrix

A $(0,1)$-matrix (also zero-one or Boolean) is a rectangular matrix for which each element of the matrix has the value of either 0 or 1. E.g.

$$A\left(\in \{0,1\}^{3\times 4}\right) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Application Areas

$(0,1)$-Matrices arise from problems in a variety of application areas including:

Application Areas

$(0,1)$-Matrices arise from problems in a variety of application areas including:

- adjacency matrices for simple graphs, representing the connectivity relationships between vertices;

Application Areas

$(0, 1)$-Matrices arise from problems in a variety of application areas including:

- term-by-document matrices generated by binary vector space model based computational information retrieval (BVIR) algorithms and applications such as search engines; an element w_{ij} in a weight matrix $W \in \{0, 1\}^{n \times m}$ is set to 1 if a term t_j appears in document D_i, with 0 otherwise;

Application Areas

$(0, 1)$-Matrices arise from problems in a variety of application areas including:

- matrix calculus applications in statistics and econometrics which generate special (0,1)-matrices such as selection, permutation, commutation, elimination, duplication, and shifting matrices;

General Matrix Decomposition

In fact, any general matrix may be decomposed into the linear combination of conformal $(0, 1)$-matrices.

General Matrix Decomposition

In fact, any general matrix may be decomposed into the linear combination of conformal $(0,1)$-matrices. For a general real matrix $A \in \mathbb{R}^{m\times n}$,

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{ij} E_{ij}, \left\{ E_{ij} = \mathbf{e}_i \mathbf{e}_j^T, _{1\le i\le m; 1\le j\le m} \right\}$$

General Matrix Decomposition

In fact, any general matrix may be decomposed into the linear combination of conformal $(0,1)$-matrices. For a general real matrix $A \in \mathbb{R}^{m\times n}$,

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{ij} E_{ij}, \left\{E_{ij} = \mathbf{e}_i \mathbf{e}_j^T,_{1\le i\le m; 1\le j\le m}\right\}$$

E.g.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Matrix-Vector Product

A matrix-vector product operation may be represented by the equation

$$Ax = y$$

where the components are conformal:
$A \in \{0,1\}^{m \times n}$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$.
The vectors may actually be over any algebraic ring for which addition and multiplication is closed and well defined.

Matrix-Vector Product

E.g.

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} e \\ f \\ g \end{pmatrix}$$

Matrix-Vector Product

The general matrix-vector product

- is an example of *level 2* operation

Matrix-Vector Product

The general matrix-vector product

- is an example of *level 2* operation
- exhibits a quadratic complexity, $O(n^2)$

Matrix-Vector Product

The general matrix-vector product

- is an example of *level 2* operation
- exhibits a quadratic complexity, $O(n^2)$
- may be blocked for modest performance gains

Matrix-Vector Product

The general matrix-vector product

- is an example of *level 2* operation
- exhibits a quadratic complexity, $O(n^2)$
- may be blocked for modest performance gains
- admits no Strassen method type decomposition

Matrix-Vector Product

The general matrix-vector product

- is an example of *level 2* operation
- exhibits a quadratic complexity, $O(n^2)$
- may be blocked for modest performance gains
- admits no Strassen method type decomposition
- $\exists < O(n^2)$ algorithms for certain patterned, low rank, and low displacement rank matrices

Matrix-Vector Product

The general matrix-vector product

- is an example of *level 2* operation
- exhibits a quadratic complexity, $O(n^2)$
- may be blocked for modest performance gains
- admits no Strassen method type decomposition
- $\exists < O(n^2)$ algorithms for certain patterned, low rank, and low displacement rank matrices

Gray Codes

A Grey code

- is a circular ordering of all 2^n binary strings of length n in which adjacent strings differ in exactly one bit.

Gray Codes

A Grey code

- is a circular ordering of all 2^n binary strings of length n in which adjacent strings differ in exactly one bit.
- All adjacent strings have a *Hamming distance* of unity.

Gray Codes

A Grey code

- is a circular ordering of all 2^n binary strings of length n in which adjacent strings differ in exactly one bit.
- All adjacent strings have a *Hamming distance* of unity.
- The Hamming distance between two bit strings is the count of differing corresponding bits.

Gray Codes

A Grey code

- One can label the vertices of a *hypercube* with bitstrings such that all *adjacent* neighbors have a bitstring differing in exactly one bit.

Gray Codes

A Grey code

- One can label the vertices of a *hypercube* with bitstrings such that all *adjacent* neighbors have a bitstring differing in exactly one bit.
- Each Gray code corresponds to a *Hamiltonian cycle* about the vertices of a hypercube of a dimension corresponding to the string length.

Differencing Method

$Ax = y$, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$, is computed via

$$y_i = \sum_{j=1}^{n} \alpha_{ij} x_j, 1 \le i \le m$$

$O(mn)$ additions and multiplications.

Differencing Method

$Ax = y$, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$, is computed via

$$y_i = \sum_{j=1}^{n} \alpha_{ij} x_j, 1 \leq i \leq m$$

$O(mn)$ additions and multiplications.
A doubly nested loop is required having one of two possible orderings.

Differencing Method

$Ax = y$, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$, is computed via

$$y_i = \sum_{j=1}^{n} \alpha_{ij} x_j, 1 \leq i \leq m$$

$O(mn)$ additions and multiplications.
A doubly nested loop is required having one of two possible orderings.

We'll select the ordering which performs an inner product ($v^T u = \alpha$)

Differencing Method

For a $(0, 1)$-matrix though, $\alpha_{ij} \in \{0, 1\}, \forall i, j$.

- either the product $\alpha_{ij} x_j$ contributes to the sum as x_j when $\alpha_{ij} = 1$

Differencing Method

For a $(0,1)$-matrix though, $\alpha_{ij} \in \{0,1\}, \forall i, j$.

- either the product $\alpha_{ij} x_j$ contributes to the sum as x_j when $\alpha_{ij} = 1$
- or it is skipped, when $\alpha_{ij} = 0$.

Differencing Method

For a $(0, 1)$-matrix though, $\alpha_{ij} \in \{0, 1\}, \forall i, j$.

- either the product $\alpha_{ij} x_j$ contributes to the sum as x_j when $\alpha_{ij} = 1$
- or it is skipped, when $\alpha_{ij} = 0$.
- $\Rightarrow$ All multiplications have been obviated leaving $\leq mn$ additions.

Differencing Method

Consider computing the difference between two elements of y, y_i and y_k:

$$\begin{aligned} y_k - y_i &= \sum_{j=1}^{n} \alpha_{kj} x_j - \sum_{j=1}^{n} \alpha_{ij} x_j \\ &= \sum_{j=1}^{n} \left(\alpha_{kj} x_j - \alpha_{ij} x_j \right) \\ &= \sum_{j=1}^{n} x_j \left(\alpha_{kj} - \alpha_{ij} \right) \end{aligned}$$

Differencing Method

However, if y_i has already been computed,

$$
\begin{aligned}
y_k &= y_i + \sum_{j=1}^{n} x_j \left(\alpha_{kj} - \alpha_{ij}\right) \\
&= y_i + \sum_{j=1}^{n} x_j \left(d_j\right), d_j = \alpha_{kj} - \alpha_{ij}
\end{aligned}
$$

Differencing Method

However, if y_i has already been computed,

$$\begin{aligned} y_k &= y_i + \sum_{j=1}^{n} x_j \left(\alpha_{kj} - \alpha_{ij}\right) \\ &= y_i + \sum_{j=1}^{n} x_j \left(d_j\right), d_j = \alpha_{kj} - \alpha_{ij} \end{aligned}$$

$\Rightarrow$ each subsequent element of y can be computed as the sum of a previously computed element with the inner product of x and the difference vector of the two rows of A, d.

Differencing Method

Then we have saved $\|A_k\|_1 - \|d\|_1$ operations in computing y_k, where $\|d\|_1$ is the *Hamming distance* between two rows of A.

Differencing Method

Then we have saved $\|A_k\|_1 - \|d\|_1$ operations in computing y_k, where $\|d\|_1$ is the *Hamming distance* between two rows of A.

For a duplicated row in A, load its previously computed value at the expense of no additions.

Differencing Method

If the remaining rows are unique, what's the lowest number of additions that we need?

Differencing Method

If the remaining rows are unique, what's the lowest number of additions that we need?
There must be a Hamming distance of at least one between each row of A and each other row.

Differencing Method

If the remaining rows are unique, what's the lowest number of additions that we need? There must be a Hamming distance of at least one between each row of A and each other row. The optimum in this case would be for there to be, for every row of A, some other row of A of unit Hamming distance. We have already defined a sequence of Boolean vectors which satisfies this condition,

Differencing Method

If the remaining rows are unique, what's the lowest number of additions that we need? There must be a Hamming distance of at least one between each row of A and each other row. The optimum in this case would be for there to be, for every row of A, some other row of A of unit Hamming distance. We have already defined a sequence of Boolean vectors which satisfies this condition,

a Gray code.

Differencing Method

- A Gray code q bits wide consists of a sequence of $r = 2^q$ bitstrings.

Differencing Method

- A Gray code q bits wide consists of a sequence of $r = 2^q$ bitstrings.
- Using the differencing method above, a $r \times q$ Gray code matrix-vector product can be computed in a maximum of r additions.

Differencing Method

- A Gray code q bits wide consists of a sequence of $r = 2^q$ bitstrings.
- Using the differencing method above, a $r \times q$ Gray code matrix-vector product can be computed in a maximum of r additions.
- If one substitutes a load for an addition when there are at most 1 nonzero bits in the row, the maximum becomes $r - q - 1$ required additions.

Differencing Method

If we have a general $(0,1)$-matrix matrix having q columns and any number of rows,

- we can first precompute a scratch y using the Gray code Matrix vector product with x.

Differencing Method

If we have a general $(0, 1)$-matrix matrix having q columns and any number of rows,

- we can first precompute a scratch y using the Gray code Matrix vector product with x.
- Then, for each row of the general matrix A, we load into y the entry in the scratch y corresponding to the bitpattern for that row.

Differencing Method

If we have a general $(0,1)$-matrix matrix having q columns and any number of rows,

- we can first precompute a scratch y using the Gray code Matrix vector product with x.
- Then, for each row of the general matrix A, we load into y the entry in the scratch y corresponding to the bitpattern for that row.
- The only additions are from the precomputation of the scratch y from the Gray code matrix. If $m > r$, we have a greater reduction in the maximum number of additions for the Gray code matrix itself.

Differencing Method

If $n > q$, one of two conditions can hold:

- $n \bmod q = 0$ in which case we will use n/q Gray code matrix stripes across A with a subsequent sum across the stripes for all m.

Differencing Method

If $n > q$, one of two conditions can hold:

- $n \bmod q = 0$ in which case we will use n/q Gray code matrix stripes across A with a subsequent sum across the stripes for all m.
- $n \bmod q \neq 0$ in which case we will use $\lfloor n/q \rfloor$ Gray code matrix stripes of width q, and one additional Gray code matrix stripe of width $n \bmod q$ with a subsequent sum across the stripes for all m.

Differencing Method

If we define $w(q, m, n)$ as the number of additions required for a general $(0, 1)$-matrix. Then for an arbitrary m and n, ignoring the $q + 1$ savings from above,

$$w(q, m, n) = \begin{cases} 2^n & \text{if } n < q \\ 2^q & \text{if } n = q \\ \frac{n}{q}\left(2^q + m\right) - m & \text{if } n > q \text{ and } 0 = n \bmod q \\ \lfloor \frac{n}{q} \rfloor 2^q + 2^{n \bmod q} + m\left(\lceil \frac{n}{q} \rceil - 1\right) & \text{if } n > q \text{ and } 0 \neq n \bmod q \end{cases}$$

The minimizing value of q may be determined informally by selecting the smallest w for a reasonably small set of q's clustered around $\lg m$.

Differencing Method

Formally, we can evaluate $\frac{\partial w}{\partial q}$ and find where it has the value of zero. For the third condition:

$$\frac{\partial w}{\partial q} = \frac{n2^q \ln 2}{q} - \frac{n(2^q + m)}{q^2}$$

$0 = \frac{\partial w}{\partial q}$ at

$$q = \frac{\mathit{LambertW}(me^{-1}) + 1}{\ln 2}$$

where $\mathit{LambertW}(x)e^{\mathit{LambertW}(x)} = x$

Exploiting $(0,1)$-Mv Products

The initial linear combination equation has little practical value unless one considers the special case where the number of distinct α_{ij} terms is smaller than the smallest of the two indices.

Exploiting $(0, 1)$-Mv Products

The initial linear combination equation has little practical value unless one considers the special case where the number of distinct α_{ij} terms is smaller than the smallest of the two indices.

Then all of the E_{ij} matrices will be added together forming denser $(0, 1)$-matrices for each set of identical α_{ij} terms.

Exploiting $(0, 1)$-Mv Products

We then get the equation

$$A = \sum_{i=1}^{k} \beta_i B_i$$

Where there are k distinct entries, β_i in A and each B_i represents a distinct $(0, 1)$-matrix.

Exploiting $(0, 1)$-Mv Products

For a single β, the matrix vector product can be written as:

$$Ax = y = \beta Bx = B(\beta x),$$

The scalar product has been reduced from $O(mn)$ to $O(n)$.

Exploiting $(0,1)$-Mv Products

For a single β, the matrix vector product can be written as:

$$Ax = y = \beta Bx = B(\beta x),$$

The scalar product has been reduced from $O(mn)$ to $O(n)$.
The previous two equations may be combined:

$$Ax = y = \left(\sum_{i=1}^{k} \beta_i B_i\right) x = \sum_{i=1}^{k} B_i(\beta_i x).$$

Exploiting $(0, 1)$-Mv Products

Lets consider the case where $k = 2$.

- The elements of the matrix A are either of only two values, $\alpha_{ij} \in \{\beta_1, \beta_2\}$.

Exploiting $(0, 1)$-Mv Products

Lets consider the case where $k = 2$.

- The elements of the matrix A are either of only two values, $\alpha_{ij} \in \{\beta_1, \beta_2\}$.
- The cardinality of the set that α_{ij} is drawn from is the same as that of the $(0, 1)$-matrix.

Exploiting $(0, 1)$-Mv Products

Lets consider the case where $k = 2$.

- The elements of the matrix A are either of only two values, $\alpha_{ij} \in \{\beta_1, \beta_2\}$.
- The cardinality of the set that α_{ij} is drawn from is the same as that of the $(0, 1)$-matrix.
- Rather than use a pair of $(0, 1)$-matrices, we can use a single one if we first affinely transform the equation with an appropriate scaling and translation. The elements of A can be scaled simply by multiplying by a scaling factor: γA getting $\gamma Ax = \gamma y$.

Exploiting $(0, 1)$-Mv Products

To translate the values of A, we need to create a conformal translation matrix, $T = \{1\}^{m\times n}$, which will serve as a basis for the uniform translation of all the entries of A.
A unit translation of A towards $+\infty$ is represented by the sum $A + T$.
We get $(A + T)x = Ax + Tx = y^{(A)} + y^{(T)}$,
where $y_j^{(T)} = \sum_{i=1}^{n} x_i$.

Exploiting $(0, 1)$-Mv Products

Translating the entries of A an arbitrary distance δ. We then augment the above equation getting: $(A + \delta T)x = Ax + \delta Tx = y^{(A)} + \delta y^{(T)}$.

Exploiting $(0, 1)$-Mv Products

Translating the entries of A an arbitrary distance δ. We then augment the above equation getting:
$(A + \delta T)x = Ax + \delta Tx = y^{(A)} + \delta y^{(T)}$.
The above equations may now be combined to represent the full affine transformation:

$$\gamma(A + \delta T)x = \gamma(Ax + \delta Tx)$$

$$= \gamma y^{(A)} + \gamma \delta y^{(T)} = \gamma(y^{(A)} + \delta y^{(T)}))$$

Exploiting $(0,1)$-Mv Products

This equation can be used to transform any two-valued matrix vector product into a $(0,1)$-matrix based vector product.

Exploiting $(0,1)$-Mv Products

This equation can be used to transform any two-valued matrix vector product into a $(0,1)$-matrix based vector product.
E.g., a common two valued A is $\{-1,1\}^{m\times n}$.

Exploiting $(0, 1)$-Mv Products

This equation can be used to transform any two-valued matrix vector product into a $(0, 1)$-matrix based vector product.
E.g., a common two valued A is $\{-1, 1\}^{m \times n}$.

This can be transformed into a $(0, 1)$-matrix based vector product by setting $\gamma = 2$ and $\delta = -\frac{1}{2}$:

Exploiting $(0,1)$-Mv Products

$$
\begin{aligned}
\{-1,1\}^{m\times n}x &= \gamma\left(\{0,1\}^{m\times n} + \delta\{1\}^{m\times n}\right)x \\
&= \gamma(A+\delta T)x \\
&= \gamma(Ax+\delta Tx) \\
&= \gamma(y^{(A)}+\delta y^{(T)})) \\
&= 2(y^{(A)}-\frac{1}{2}y^{(T)})) \\
&= 2y^{(A)}-y^{(T)})).
\end{aligned}
$$

Conclusion

- We were able to demonstrate a differencing method for reducing the number of arithmetic operations within a $(0, 1)$-matrix vector product.

Conclusion

- We were able to demonstrate a differencing method for reducing the number of arithmetic operations within a $(0, 1)$-matrix vector product.
- We used that method to more efficiently perform Gray code matrix vector products.

Conclusion

- Using one or more Gray code matrix vector products, we showed how to reduce the number of arithmetic operations for a general $(0,1)$-matrix vector product.

Conclusion

- Using one or more Gray code matrix vector products, we showed how to reduce the number of arithmetic operations for a general $(0,1)$-matrix vector product.
- We then showed how $(0,1)$-matrix vector products may be applied to the performance of certain restricted classes of more general matrix vector products.