Speeding up OMP

In the last lecture note, we introduced orthogonal matching pursuit, an algorithm that can extract information from sparse signals. Recall that in each iteration of the algorithm, we need to compute the projection of the measurement vector $\vec{y}$ onto the subspace spanned by the signatures computed so far. If we let A_j be a matrix whose columns are the signatures found so far in iteration j and let $\vec{y}$ be the measurement vector, we compute $A_j\left(A_j^T A_j\right)^{-1} A_j^T \vec{y}$ to get the projection of $\vec{y}$ onto span(A_j). Matrix inversion and matrix multiplication can be computationally expensive. Is there a way to avoid doing such computations? Yes! It turns out that if the columns of A_j are mutually orthogonal to each other, the projection of $\vec{y}$ onto span(A_j) is the sum of the projection of $\vec{y}$ onto each of the columns of A_j. Recall that the projection of a vector $\vec{y}$ on any other nonzero vector $\vec{b}$ of the same size is

$$\vec{y}_b = \frac{\vec{y}^T\vec{b}}{\left\|\vec{b}\right\|^2}. \tag{1}$$

This is fast to compute as computing the dot product of two vectors and the norm of a vector takes linear in time in the number of components in the vectors.

Let's take a look at the case where $j = 2$ and the signatures are mutually orthogonal. Suppose the signatures found so far are $\vec{v}_1$ and $\vec{v}_2$, i.e., $A_2 = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix}$. Since $\vec{v}_1$ and $\vec{v}_2$ are orthogonal to each other, we have $\vec{v}_1{}^T\vec{v}_2 = 0$. The projection of $\vec{y}$ on A_2 is

$$A_2\left(A_2^T A_2\right)^{-1} A_2^T \vec{y}. \tag{2}$$

Let's first compute the term $\left(A_2^T A_2\right)^{-1}$:

$$\begin{aligned}
A_2^T A_2 &= \begin{bmatrix} — & \vec{v}_1{}^T & — \\ — & \vec{v}_2{}^T & — \end{bmatrix} \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} && (3)\\
&= \begin{bmatrix} \vec{v}_1{}^T\vec{v}_1 & \vec{v}_1{}^T\vec{v}_2 \\ \vec{v}_2{}^T\vec{v}_1 & \vec{v}_2{}^T\vec{v}_2 \end{bmatrix} && (4)\\
&= \begin{bmatrix} \|\vec{v}_1\|^2 & 0 \\ 0 & \|\vec{v}_2\|^2 \end{bmatrix}. && (5)\\
& && (6)
\end{aligned}$$

Thus, we have

$$\left(A_2^T A_2\right)^{-1} = \begin{bmatrix} \frac{1}{\|\vec{v}_1\|^2} & 0 \\ 0 & \frac{1}{\|\vec{v}_2\|^2} \end{bmatrix}. \tag{7}$$

Then the projection of $\vec{y}$ onto span(A_2) is

$$\vec{y}_{A_2} = A_2 \left(A_2^T A_2\right)^{-1} A_2^T \vec{y} \tag{8}$$

$$= \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} \begin{bmatrix} \frac{1}{\|\vec{v}_1\|^2} & 0 \\ 0 & \frac{1}{\|\vec{v}_2\|^2} \end{bmatrix} \begin{bmatrix} — & \vec{v}_1{}^T & — \\ — & \vec{v}_2{}^T & — \end{bmatrix} \vec{y} \tag{9}$$

$$= \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} \begin{bmatrix} \frac{1}{\|\vec{v}_1\|^2} & 0 \\ 0 & \frac{1}{\|\vec{v}_2\|^2} \end{bmatrix} \begin{bmatrix} \vec{v}_1{}^T \vec{y} \\ \vec{v}_2{}^T \vec{y} \end{bmatrix} \tag{10}$$

$$= \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} \begin{bmatrix} \frac{\vec{v}_1{}^T \vec{y}}{\|\vec{v}_1\|^2} \\ \frac{\vec{v}_2{}^T \vec{y}}{\|\vec{v}_2\|^2} \end{bmatrix} \tag{11}$$

$$= \left(\frac{\vec{v}_1{}^T \vec{y}}{\|\vec{v}_1\|^2}\right) \vec{v}_1 + \left(\frac{\vec{v}_2{}^T \vec{y}}{\|\vec{v}_2\|^2}\right) \vec{v}_2. \tag{12}$$

Observe that the first term in the sum above is the projection of $\vec{y}$ onto $\vec{v}_1$ and the second term is the projection of $\vec{y}$ onto $\vec{v}_2$. Generalizing, the projection of $\vec{y}$ onto span(A_n) where A_n has mutually orthogonal columns is

$$\vec{y}_{A_n} = \left(\frac{\vec{v}_1{}^T \vec{y}}{\|\vec{v}_1\|^2}\right) \vec{v}_1 + \left(\frac{\vec{v}_2{}^T \vec{y}}{\|\vec{v}_2\|^2}\right) \vec{v}_2 + \cdots + \left(\frac{\vec{v}_n{}^T \vec{y}}{\|\vec{v}_n\|^2}\right) \vec{v}_n. \tag{13}$$

Furthermore, observe that if $\vec{v}_1, \ldots, \vec{v}_n$ are unit vectors, i.e., they all have length 1, the above further reduces to

$$\vec{y}_{A_n} = \left(\vec{v}_1{}^T \vec{y}\right) \vec{v}_1 + \left(\vec{v}_2{}^T \vec{y}\right) \vec{v}_2 + \cdots + \left(\vec{v}_n{}^T \vec{y}\right) \vec{v}_n, \tag{14}$$

which further reduces our computation.

We see that we can speed up OMP considerably if the signatures $\vec{v}_1, \ldots, \vec{v}_n$ are mutually orthogonal to each other and are of unit length. Now the question is how do we convert any set of linearly independent vectors to a set of mutually orthogonal unit vectors that span the same vector space? We will answer this in the next section.

Gram Schmidt Process

Before we begin, let's remind ourselves that the following subspaces are equivalent for any pairs of linearly independent vectors $\vec{v}_1, \vec{v}_2$:

- span$(\vec{v}_1,\, \vec{v}_2)$
- span$(\vec{v}_1,\, \alpha\vec{v}_2)$
- span$(\vec{v}_1,\, \vec{v}_1 + \vec{v}_2)$
- span$(\vec{v}_1,\, \vec{v}_1 - \vec{v}_2)$
- span$(\vec{v}_1,\, \vec{v}_2 - \alpha\vec{v}_1)$

Now what should α be if we would like $\vec{v}_1$ and $\vec{v}_2 - \alpha\vec{v}_1$ to be orthogonal to each other? Intuitively, $\alpha\vec{v}_1$ should be the projection of $\vec{v}_2$ onto $\vec{v}_1$. Let's solve this algebraically using the definition of orthogonality:

$$\vec{v}_1 \text{ and } \vec{v}_2 - \alpha\vec{v}_1 \text{ are orthogonal} \tag{15}$$
$$\Leftrightarrow \vec{v}_1^T(\vec{v}_2 - \alpha\vec{v}_1) = 0 \tag{16}$$
$$\Leftrightarrow \vec{v}_1^T\vec{v}_2 - \alpha\|\vec{v}_1\|^2 = 0 \tag{17}$$
$$\Leftrightarrow \alpha = \frac{\vec{v}_1^T\vec{v}_2}{\|\vec{v}_1\|^2} \tag{18}$$

Definition 19.1 (Orthonormal): A set of vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is orthonormal if all the vectors are mutually orthogonal to each other and all are of unit length.

Gram Schmidt is an algorithm that takes a set of linearly independent vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ and generates an orthonormal set of vectors $\{w_1, \ldots, w_n\}$ that span the same vector space as the original set. Concretely, $\{w_1, \ldots, w_n\}$ needs to satisfy the following:

- span($\{v_1, \ldots, v_n\}$) = span($\{w_1, \ldots, w_n\}$)
- $\{w_1, \ldots, w_n\}$ is an orthonormal set of vectors

Now let's see how we can do this with a set of three vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ that is linearly independent of each other.

- **Step 1:** Find unit vector $\vec{w}_1$ such that span($\{\vec{w}_1\}$) = span($\{\vec{v}_1\}$).
  Since span($\{\vec{v}_1\}$) is a one dimensional vector space, the unit vector that span the same vector space would just be the normalized vector point at the same direction as $\vec{v}_1$. We have

  $$\vec{w}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}. \tag{19}$$

- **Step 2:** Given $\vec{w}_1$ from the previous step, find $\vec{w}_2$ such that $span(\{\vec{w}_1, \vec{w}_2\}) = span(\{\vec{v}_1, \vec{v}_2\})$ and orthogonal to $\vec{w}_1$. We know that $\vec{v}_2-$ (the projection of $\vec{v}_2$ on $\vec{w}_1$) would be orthogonal to $\vec{w}_1$ as seen earlier. Hence, a vector $\vec{e}_2$ orthogonal to $\vec{w}_1$ where $span(\{\vec{w}_1, \vec{e}_2\}) = span(\{\vec{v}_1, \vec{v}_2\})$ is

  $$\vec{e}_2 = \vec{v}_2 - \left(\vec{v}_2^T\vec{w}_1\right)\vec{w}_1. \tag{20}$$

  Normalizing, we have $\vec{w}_2 = \frac{\vec{e}_2}{\|\vec{e}_2\|}$.

- **Step 3:** Now given $\vec{w}_1$ and $\vec{w}_2$ in the previous steps, we would like to find $\vec{w}_3$ such that $span(\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}) = span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\})$. We know that the projection of $\vec{v}_3$ onto the subspace spanned by $\vec{w}_1, \vec{w}_2$ is

  $$\left(\vec{v}_3^T\vec{w}_2\right)\vec{w}_2 + \left(\vec{v}_3^T\vec{w}_1\right)\vec{w}_1. \tag{21}$$

  We know that

  $$\vec{e}_3 = \vec{v}_3 - \left(\vec{v}_3^T\vec{w}_2\right)\vec{w}_2 - \left(\vec{v}_3^T\vec{w}_1\right)\vec{w}_1 \tag{22}$$

  is orthogonal to $\vec{w}_1$ and $\vec{w}_2$. Normalizing, we have $\vec{w}_3 = \frac{\vec{e}_3}{\|\vec{e}_3\|}$.

We can generalize the above procedure for any number of linearly independent vectors as follows:

Inputs:

- A set of linearly independent vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$.

Outputs:

- An orthonormal set of vectors $\{\vec{w}_1, \ldots, \vec{w}_n\}$ where span($\{\vec{v}_1, \ldots, \vec{v}_n\}$) = span($\{\vec{w}_1, \ldots, \vec{w}_n\}$).

procedure GRAM SCHMIDT($\vec{v}_1, \ldots, \vec{v}_n$)
    $\vec{w}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}$
    for $i = 2 \ldots n$ **do**
        $\vec{e}_i \leftarrow \vec{v}_i - \sum_{j=1}^{i-1} \left(\vec{v}_i^T \vec{e}_j\right) \vec{w}_j$
        $\vec{w}_i \leftarrow \frac{\vec{e}_i}{\|\vec{e}_i\|}$
    end for
end procedure