Gram Schmidt Algorithm

Takes a set of linearly independent vectors $\{\ v_1, ..., \ v_n\}$, generates orthonormal set of vectors $\{w_1, ..., w_n\}$ that span the same vector space as the original set. $\{w_1, ..., w_n\}$ needs to satisfy the following:

- $\text{span}(\{v_1, ..., v_n\}) = \text{span}(\{w1, ..., wn\})$
- $\{w_1, ..., w_n\}$ is an orthonormal set of vectors.

Step 1: $\vec{w}_1 = \frac{\vec{v}_1}{\|v_1\|}$ **Step 2:** $e_2 = \vec{v}_2 - (\vec{v}_2^T\vec{w}_1)\vec{w}_1$; $\vec{w}_2 = \frac{\vec{v}_2}{\|v_2\|}$; **Step 3:** $e3 = \vec{v}_3 - (\vec{v}_3^T\vec{w}_2)\vec{w}_2 - (\vec{v}_3^T\vec{w}1)\vec{w}_1$; $\vec{w}_3 = \frac{\vec{v}_3}{\|v_3\|}$ **Continue summation for further vectors**

Cauchy-Schwarz: $\|\vec{x}, \vec{y}\| \leq \|\vec{x}\| \cdot \|\vec{y}\|$ **Triangle Inequality:** $\|\vec{x} + \vec{y}\| \leq \|x\| + \|y\|$
Inner/Dot product: $\langle \vec{x}, \vec{y} \rangle = \vec{x}^T\vec{y}$
If x, y, z are vectors and c is a scalar then, $\langle x, y\rangle = \langle y, x\rangle$ (symmetry); $\langle cx, y\rangle = c\langle x, y\rangle$ (homogeniety); $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ (additivity) In $R^2 : \langle x, y\rangle = |x||y|cos(\theta)$ If x orthogonal to y, $\langle x, y\rangle = 0$; Thus, **orthogonality** in R^n = inner product of 0 **Outerproduct:** $\vec{x} \otimes \vec{y} = \vec{x}\vec{y}^T$ The order of the vectors does not matter when you take an inner product, but it does matter when you take an outer product.
The larger the magnitude of the inner/dot product, the more similar the two vectors are.

Linear Least Squares: Step 1: $\hat{x} = (\vec{a}^T\vec{a})^{-1}\vec{a}^T b$ **Step 2:** $e = \vec{b} - \vec{\hat{b}} = \vec{b} - \hat{x}\vec{a}$ **Sum of Square Errors:** $e^T\vec{e}$
LLSE with Matrix A: $\hat{x} = (A^TA)^{-1}A^T\vec{b}$ **Euclidean Norm:** $\|\vec{x}\| = \sqrt{\vec{x}_1^2 + \vec{x}_2^2 + ... + \vec{x}_n^2} = \sqrt{\langle \vec{x}, \vec{x}\rangle}$
Projection: Project vector a onto vector b: $\text{proj}_b a = \frac{\langle a,b\rangle \cdot b}{\langle b,b\rangle}$ **Direction of projection:** $\hat{b} = \frac{b}{|b|} = \frac{b}{\sqrt{(\langle b,b\rangle)}}$ **Magnitude of projection:** $|a|\cos(\theta) = |a||\hat{b}|cos(\theta) = \langle a, \hat{b}\rangle$ Thus, projection vector is: $\langle a, \hat{b}\rangle\hat{b} = \frac{\langle a,b\rangle \cdot b}{\langle b,b\rangle}$
OMP: Suppose $\vec{x} \in R^d$ is the sparse vector we want to measure, with m non-zero entries (its sparsity). Let S be the nd measurement matrix with columns $\vec{S}_i$ such that: $\begin{bmatrix}\vec{S}_1 & \cdots & \vec{S}_d\end{bmatrix}\vec{x} = \vec{y}$ where the vector $\vec{y} \in R^n$ is the results from the n measurements of $\vec{x}$. **OMP works even when the number of measurements is less than the number of dimensions ($n < d$) Procedure:** 1. Find the vector $\vec{S}^*$ (not chosen before) with the highest correlation with $\vec{y}$. 2. Use least squares to project $\vec{y}$ onto the subspace spanned by the vectors already found in previous steps and $\vec{S}^*$, then find the residue r (this is the quantity minimized by least squares). 3. Repeat the above steps using r in place of $\vec{y}$ for a total of m times, the sparsity of vector $\vec{x}$.

12. One Magical Procedure (27+5 points)

Suppose we have a vector $\vec{x} \in \mathbb{R}^5$ and an $n \times 5$ measurement matrix M defined by column vectors $\vec{c}_1, \cdots, \vec{c}_5$ such that:

$$M\vec{x} = \begin{bmatrix} | & & | \\ \vec{c}_1 & \cdots & \vec{c}_5 \\ | & & | \end{bmatrix} \vec{x} \approx \vec{b}$$

We can treat the vector $\vec{b} \in \mathbb{R}^n$ as a noisy measurement of the vector $\vec{x}$, with measurement matrix M and some additional noise in it as well.

You also know that the **true $\vec{x}$ is sparse — it only has two non-zero entries and all the rest of the entries are zero in reality.** Our goal is to recover this original $\vec{x}$ as best we can.

However, your intern has managed to lose not only the measurements $\vec{b}$ but the entire measurement matrix M as well!

Fortunately, you have found a backup in which you have all the pairwise inner-products $\langle \vec{c}_i, \vec{c}_j\rangle$ between the columns of M and each other, as well as all the inner products $\left\langle \vec{c}_i, \vec{b}\right\rangle$ between the columns of M and the vector $\vec{b}$. Finally, you also find the inner-product $\left\langle \vec{b}, \vec{b}\right\rangle$ of $\vec{b}$ with itself.

All the information you have is captured in the following table of inner-products. (These are not the vectors themselves.)

$\langle \cdot,\cdot \rangle$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$	$\vec{b}$
$\vec{c}_1$	2	0	1	-1	1	1
$\vec{c}_2$		2	1	-1	-1	-5
$\vec{c}_3$			2	0	-1	2
$\vec{c}_4$				2	-1	6
$\vec{c}_5$					2	-1
$\vec{b}$						29

(So, for example, if you read this table, you will see that the inner product $\langle \vec{c}_2, \vec{c}_3\rangle = 1$, the inner product $\left\langle \vec{c}_3, \vec{b}\right\rangle = 2$, and that the inner product $\left\langle \vec{b}, \vec{b}\right\rangle = 29$. By symmetry of the real inner product, $\langle \vec{c}_3, \vec{c}_2\rangle = 1$ as well.)

Your goal is to find which entries of $\vec{x}$ are non-zero, and what their values are.

(a) (4 points) **Use the information in the table above to answer which of the $\vec{c}_1, \ldots, \vec{c}_5$ has the largest magnitude inner product with $\vec{b}$?**

c_4

(b) (5 points) Let the vector with the largest magnitude inner product with $\vec{b}$ be $\vec{c}_a$. Let $\vec{b}_p$ be the projection of $\vec{b}$ onto $\vec{c}_a$. **Write $\vec{b}_p$ symbolically as an expression only involving $\vec{c}_a$, $\vec{b}$ and their inner-products with themselves and each other.**

$\vec{b}_p = \frac{\langle \vec{c}_a, \vec{b}\rangle}{\langle c_a, c_a\rangle} c_a$

(c) (10 points) **Use the information in the table above to find which of the column vectors $\vec{c}_1, \ldots, \vec{c}_5$ has the largest magnitude inner product with the residue $\vec{b} - \vec{b}_p$.**
(Hint: the linearity of inner products might prove useful.)

$\vec{b} - \vec{b}_p = \vec{b} - \frac{\langle \vec{c}_a, \vec{b}\rangle}{\langle c_a, c_a\rangle} c_a$

Let c_b be a prospective column vector not c_a

$\langle c_b, \vec{b} - \vec{b}_p\rangle =$

$\langle c_b, \vec{b} - \frac{\langle \vec{c}_a, \vec{b}\rangle}{\langle c_a, c_a\rangle} c_a\rangle =$

$\langle c_b, \vec{b}\rangle - \langle c_b, \frac{\langle \vec{c}_a, \vec{b}\rangle}{\langle c_a, c_a\rangle} c_a\rangle =$

$\langle c_b, \vec{b}\rangle - \langle c_b, c_a\rangle \cdot \frac{\langle c_a, b\rangle}{\langle c_a, c_a\rangle}$

filling values in:

$\langle c_b, \vec{b}\rangle - \langle c_b, c_4\rangle \cdot \frac{6}{2}$

	$\langle c_b, b\rangle$	$-3\langle c_b, c_4\rangle$	sum
c_1	1	3	4
c_2	-5	3	-2
c_3	2	0	2
c_5	-1	3	2

$\boxed{c_1; \ c_1 \text{ has the greatest magnitude value of } \langle c_b, \vec{b}\rangle - 3\langle c_b, c_4\rangle}$

(d) (8 points) Suppose the vectors we found in parts (a) and (c) are $\vec{c}_a$ and $\vec{c}_c$. These correspond to the components of $\vec{x}$ that are non-zero, that is, $\vec{b} \approx x_a\vec{c}_a + x_c\vec{c}_c$. However, there might be noise in the measurements $\vec{b}$, so we want to find the linear-least-squares estimates $\hat{x}_a$ and $\hat{x}_c$. **Write a matrix expression for $\begin{bmatrix}\hat{x}_a \\ \hat{x}_c\end{bmatrix}$ in terms of appropriate matrices filled with the inner products of $\vec{c}_a, \vec{c}_c, \vec{b}$.**

$$A^T A \vec{x} = A^T b \qquad A = \begin{bmatrix}\vec{c}_a & \vec{c}_c\end{bmatrix}$$

$$A^T A = \begin{bmatrix}\vec{c}_a^{\,T} \\ \vec{c}_c^{\,T}\end{bmatrix}\begin{bmatrix}\vec{c}_a & \vec{c}_c\end{bmatrix} = \begin{bmatrix}\langle c_a, c_a\rangle, & \langle c_a, c_c\rangle \\ \langle c_a, c_c\rangle, & \langle c_c, c_c\rangle\end{bmatrix}$$

$$A^T b = \begin{bmatrix}\vec{c}_a^{\,T} \\ \vec{c}_c^{\,T}\end{bmatrix}\begin{bmatrix}b\end{bmatrix} = \begin{bmatrix}\langle c_a, b\rangle \\ \langle c_c, b\rangle\end{bmatrix}$$

$$\begin{bmatrix}\hat{x}_a \\ \hat{x}_c\end{bmatrix} = \begin{bmatrix}\langle c_a, c_a\rangle, & \langle c_a, c_c\rangle \\ \langle c_a, c_c\rangle, & \langle c_c, c_c\rangle\end{bmatrix}^{-1}\begin{bmatrix}\langle c_a, b\rangle \\ \langle c_c, b\rangle\end{bmatrix}$$

$C_a = C_p$
$C_c = C_1$

(e) (BONUS: 5 points) **Compute the numerical values of $\hat{x}_a$ and $\hat{x}_c$ using the information in the table.**

$$\begin{bmatrix}\hat{x}_a \\ \hat{x}_c\end{bmatrix} = \begin{bmatrix}2 & -1 \\ -1 & 2\end{bmatrix}^{-1}\begin{bmatrix}6 \\ 1\end{bmatrix}$$

$$\begin{bmatrix}2 & -1 & 1 & 0 \\ -1 & 2 & 0 & 1\end{bmatrix} \sim \begin{bmatrix}1 & -1/2 & 1/2 & 0 \\ -1 & 2 & 0 & 1\end{bmatrix} \sim \begin{bmatrix}1 & -1/2 & 1/2 & 0 \\ 0 & 3/2 & 1/2 & 1\end{bmatrix} \sim \begin{bmatrix}2 & -1 & 1 & 0 \\ 0 & 1 & 1/3 & 2/3\end{bmatrix} \sim$$

$$\begin{bmatrix}2 & 0 & 4/3 & 2/3 \\ 0 & 1 & 1/3 & 2/3\end{bmatrix} \sim \begin{bmatrix}1 & 0 & 2/3 & 1/3 \\ 0 & 1 & 1/3 & 2/3\end{bmatrix}$$

$$\begin{bmatrix}\hat{x}_a \\ \hat{x}_c\end{bmatrix} = \begin{bmatrix}2/3 & 1/3 \\ 1/3 & 2/3\end{bmatrix}\begin{bmatrix}6 \\ 1\end{bmatrix} = \begin{bmatrix}4 + 1/3 \\ 2 + 2/3\end{bmatrix} = \begin{bmatrix}13/3 \\ 8/3\end{bmatrix}$$

procedure OMP($\mathbf{S}$, $\vec{y}$, m)
    $\vec{r} \leftarrow \vec{y}$
    $\mathbf{A}_0 \leftarrow [\,]$
    $\Lambda_0 \leftarrow [\,]$
    for j in [1, m] **do**
        $k \leftarrow argmax_i\ corr(\vec{S}_i, \vec{r})$
        $\mathbf{A}_j \leftarrow [\mathbf{A}_{j-1}\ \vec{S}_k]$
        $\Lambda_j \leftarrow \Lambda_{j-1} \cup k$
        $\hat{\vec{x}}_j = (A_j^T A_j)^{-1} A_j^T \vec{y}$
        $\hat{\vec{y}}_j = A_j \hat{\vec{x}}_j$
        $\vec{r} \leftarrow \vec{y} - \hat{\vec{y}}$
    end for
    return $\vec{x}_m$
end procedure

Diagonlization: $A^n = V\Lambda^n V^{-1} = \sum_{i=1}^{N} \lambda_i^n \vec{v}_i \vec{w}_i^T$

Pagerank Importance Score
Converges to the value $A^n s[0]$ where:

$$\lim_{n\to\infty} A^n = \lambda_i \vec{v}_i \vec{w}_i^T = \vec{v}_i \vec{w}_i^T \text{ where } \lambda_i = 1 \text{ if } |\lambda_i| < 1 \forall i = 1, \ldots, N-1 \tag{1}$$

Determinants 1. If you scale a row/col of a matrix by α, the determinant of the matrix is multiplied by α 2. If you add a scalar multiple of a row/col to any other row/col, the determinant doesn?t change 3. If you swap rows, the determinant is multiplied by -1
Solve for Eigenvalues/Eigenvectors

Solve for λ: $\det(A - \lambda I) = 0$
Determinant of a $n \times n$ matrix
Mult. each elem. in the first row by the det. of the $n-1 \times n-1$ matrix not in that element's row or column
Take an **alternating sum** of the products from the previous step

$$\det\begin{bmatrix}a & b & c \\ d & e & f \\ g & h & i\end{bmatrix} = a \times \det\begin{bmatrix}e & i \\ f & h\end{bmatrix} - b \times \det\begin{bmatrix}d & i \\ f & g\end{bmatrix} + c \times \det\begin{bmatrix}d & h \\ e & g\end{bmatrix}$$

Plug resulting Eigenvalues λ_i into $A - \lambda_i I$: $B_i = A - \lambda_i I$
Solve for Eigenvectors $\vec{v}_i$ s.t. This will be the linear combination of the columns of B_i that cause $B_i\vec{v}_i = 0$: $(A - \lambda_i I)\vec{v}_i = B_i\vec{v}_i = 0$

Pagerank State Transition: $\vec{s}[n+1] = A\vec{s}[n] = \alpha_1\lambda_1^n\vec{v}_1 + \alpha_2\lambda_2^n\vec{v}_2$ **Cross Correlation:** $T = \text{argmax}(\text{crossCorr}(x, y)) = \text{argmax}(C_x^T y)$; **Intuition:** We shift x around until it has the maximum correlation with y, the received signal **Locationing:** With $n+1$ beacons, uniquely locate a beacon in R^n; Each beacon provides equation: $(x-a)^2 + (y-b)^2 = r^2$ (x, y = location of object, a, b = location of beacon, r = distance from beacon to object (may be noisy)); Convert nonlinear equations to one linear equation by expanding and subtracting; Convert equation to matrix A and use **LLSE:** $\hat{x} = (A^T A)^{-1} A^T \vec{b}$ to solve for x, y. **Ways to Solve Lin. Sys.:** Perfect information: $x = A^{-1}b$; Noisy information: $x = (A^T A)^{-1} A^T \vec{b}$; Sparse information: $x = \text{OMP}(A, b)$