1 Gram-Schmidt orthogonalization

- Given: a set of vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k$ in $\mathbb{R}^n$.
- Construct: as set of unit-length orthogonal vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_k$.
  - unit-length: $\|\mathbf{w}_i\| = 1$
  - orthogonal: $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ for $i \neq j$
- Tool: projection
  - $\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}$
  - $\mathbf{w} = \mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$
  - Note:

$$\begin{aligned}\mathbf{w}\cdot\mathbf{u} &= \left(\mathbf{v} - \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}\right)\cdot\mathbf{u} \\ &= \mathbf{v}\cdot\mathbf{u} - \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}\cdot\mathbf{u} \\ &= \mathbf{v}\cdot\mathbf{u} - \mathbf{u}\cdot\mathbf{v} \\ &= 0\end{aligned}$$

  - Note: if $\|\mathbf{u}\| = 1$ then $\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \mathbf{v} - (\mathbf{u}\cdot\mathbf{v})\mathbf{u}$
- Works with multiple vectors at once
  - Assume $\mathbf{u}_i \cdot \mathbf{u}_j = 0$
  - Let $\mathbf{w} = \mathbf{v} - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}) - \mathrm{proj}_{\mathbf{u}_2}(\mathbf{v}) - \mathrm{proj}_{\mathbf{u}_3}(\mathbf{v})$
  - Then: $\mathbf{w} = \mathbf{v} - \left(\frac{\mathbf{u}_1\cdot\mathbf{v}}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1 - \left(\frac{\mathbf{u}_2\cdot\mathbf{v}}{\mathbf{u}_2\cdot\mathbf{u}_2}\right)\mathbf{u}_2 - \left(\frac{\mathbf{u}_3\cdot\mathbf{v}}{\mathbf{u}_3\cdot\mathbf{u}_3}\right)\mathbf{u}_3$

$$\begin{aligned}\mathbf{w}\cdot\mathbf{u}_2 &= \left(\mathbf{v} - \left(\frac{\mathbf{u}_1\cdot\mathbf{v}}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1 - \left(\frac{\mathbf{u}_2\cdot\mathbf{v}}{\mathbf{u}_2\cdot\mathbf{u}_2}\right)\mathbf{u}_2 - \left(\frac{\mathbf{u}_3\cdot\mathbf{v}}{\mathbf{u}_3\cdot\mathbf{u}_3}\right)\mathbf{u}_3\right)\cdot\mathbf{u}_2 \\ &= \mathbf{v}\cdot\mathbf{u}_2 - \left(\frac{\mathbf{u}_1\cdot\mathbf{v}}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\underbrace{\mathbf{u}_1\cdot\mathbf{u}_2}_{0} - \left(\frac{\mathbf{u}_2\cdot\mathbf{v}}{\mathbf{u}_2\cdot\mathbf{u}_2}\right)\mathbf{u}_2\cdot\mathbf{u}_2 - \left(\frac{\mathbf{u}_3\cdot\mathbf{v}}{\mathbf{u}_3\cdot\mathbf{u}_3}\right)\underbrace{\mathbf{u}_3\cdot\mathbf{u}_2}_{0} \\ &= \mathbf{v}\cdot\mathbf{u}_2 - \mathbf{u}_2\cdot\mathbf{v} \\ &= 0\end{aligned}$$

  - Note: same simplification if $\|\mathbf{u}_i\| = 1$
- Algorithm

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{u}_1 & \mathbf{w}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \\ \mathbf{v}_2 &= \mathbf{u}_2 - (\mathbf{w}_1\cdot\mathbf{u}_2)\mathbf{w}_1 & \mathbf{w}_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} \\ \mathbf{v}_3 &= \mathbf{u}_3 - (\mathbf{w}_1\cdot\mathbf{u}_3)\mathbf{w}_1 - (\mathbf{w}_2\cdot\mathbf{u}_3)\mathbf{w}_2 & \mathbf{w}_3 &= \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} \\ \mathbf{v}_4 &= \mathbf{u}_4 - (\mathbf{w}_1\cdot\mathbf{u}_4)\mathbf{w}_1 - (\mathbf{w}_2\cdot\mathbf{u}_4)\mathbf{w}_2 - (\mathbf{w}_3\cdot\mathbf{u}_4)\mathbf{w}_3 & \mathbf{w}_4 &= \frac{\mathbf{v}_4}{\|\mathbf{v}_4\|} \\ &\ldots\end{aligned}$$

2 Relation to QR

- Rearrange the algorithm (eliminate $\mathbf{v}_i$)

$$\begin{aligned}
\|\mathbf{v}_1\|\mathbf{w}_1 &= \mathbf{u}_1 \\
\|\mathbf{v}_2\|\mathbf{w}_2 &= \mathbf{u}_2 - (\mathbf{w}_1 \cdot \mathbf{u}_2)\mathbf{w}_1 \\
\|\mathbf{v}_3\|\mathbf{w}_3 &= \mathbf{u}_3 - (\mathbf{w}_1 \cdot \mathbf{u}_3)\mathbf{w}_1 - (\mathbf{w}_2 \cdot \mathbf{u}_3)\mathbf{w}_2 \\
\|\mathbf{v}_4\|\mathbf{w}_4 &= \mathbf{u}_4 - (\mathbf{w}_1 \cdot \mathbf{u}_4)\mathbf{w}_1 - (\mathbf{w}_2 \cdot \mathbf{u}_4)\mathbf{w}_2 - (\mathbf{w}_3 \cdot \mathbf{u}_4)\mathbf{w}_3 \\
&\dots
\end{aligned}$$

- Rearrange the algorithm (put $\mathbf{w}_i$ stuff on one side)

$$\begin{aligned}
\mathbf{u}_1 &= \|\mathbf{v}_1\|\mathbf{w}_1 \\
\mathbf{u}_2 &= \|\mathbf{v}_2\|\mathbf{w}_2 + (\mathbf{w}_1 \cdot \mathbf{u}_2)\mathbf{w}_1 \\
\mathbf{u}_3 &= \|\mathbf{v}_3\|\mathbf{w}_3 + (\mathbf{w}_1 \cdot \mathbf{u}_3)\mathbf{w}_1 + (\mathbf{w}_2 \cdot \mathbf{u}_3)\mathbf{w}_2 \\
\mathbf{u}_4 &= \|\mathbf{v}_4\|\mathbf{w}_4 + (\mathbf{w}_1 \cdot \mathbf{u}_4)\mathbf{w}_1 + (\mathbf{w}_2 \cdot \mathbf{u}_4)\mathbf{w}_2 + (\mathbf{w}_3 \cdot \mathbf{u}_4)\mathbf{w}_3 \\
&\dots
\end{aligned}$$

- Give the scalars names: $r_{ii} = \|\mathbf{v}_1\|$, $r_{ij} = \mathbf{w}_i \cdot \mathbf{u}_j$ for $i < j$

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{w}_1 r_{11} \\
\mathbf{u}_2 &= \mathbf{w}_2 r_{22} + \mathbf{w}_1 r_{12} \\
\mathbf{u}_3 &= \mathbf{w}_3 r_{33} + \mathbf{w}_1 r_{13} + \mathbf{w}_2 r_{23} \\
\mathbf{u}_4 &= \mathbf{w}_4 r_{44} + \mathbf{w}_1 r_{14} + \mathbf{w}_2 r_{24} + \mathbf{w}_3 r_{34} \\
&\dots
\end{aligned}$$

- Writing as matrices

$$\underbrace{\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 & \mathbf{w}_4 \end{pmatrix}}_{\mathbf{Q}} \underbrace{\begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ & r_{22} & r_{23} & r_{24} \\ & & r_{33} & r_{34} \\ & & & r_{44} \end{pmatrix}}_{\mathbf{R}}$$

- Note: $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ since $\mathbf{w}_i \cdot \mathbf{w}_i = 1$ and $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ otherwise, by construction
- $\mathbf{A} = \mathbf{QR}$.
  - Normally $\mathbf{A}$ is square
  - What if non-square?
  - What if $\mathbf{A}$ is not full rank?

3 Modified Gram-Schmidt orthogonalization

- In practice, Gram-Schmidt tends to lose orthogonalization between vectors (numerically unstable).
- Original: $\mathbf{v}_4 = \mathbf{u}_4 - \text{proj}_{\mathbf{w}_1}(\mathbf{u}_4) - \text{proj}_{\mathbf{w}_2}(\mathbf{u}_4) - \text{proj}_{\mathbf{w}_3}(\mathbf{u}_4) \qquad \mathbf{w}_4 = \frac{\mathbf{v}_4}{\|\mathbf{v}_4\|}$
- Modified: $\mathbf{v}_2 = \mathbf{u}_4 - \text{proj}_{\mathbf{w}_1}(\mathbf{u}_4) \qquad \mathbf{v}_3 = \mathbf{v}_2 - \text{proj}_{\mathbf{w}_2}(\mathbf{v}_2) \qquad \mathbf{v}_4 = \mathbf{v}_3 - \text{proj}_{\mathbf{w}_3}(\mathbf{v}_3) \qquad \mathbf{w}_4 = \frac{\mathbf{v}_4}{\|\mathbf{v}_4\|}$
- Project in sequence, not all at once.

4 Householder QR

- Householder transform: $\mathbf{Q} = \mathbf{I} - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T$
  - Want reflection $\mathbf{Q}$ such that $\mathbf{Q}\mathbf{u} = \alpha\mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)$
  - $\mathbf{v} = \mathbf{u} - \alpha\mathbf{e}_1$, where $\alpha = -\text{sgn}(\mathbf{u}_1)\|\mathbf{u}\|$

$$
\begin{aligned}
\mathbf{v} &= \mathbf{u} - \alpha\mathbf{e}_1 \\
\mathbf{v}\cdot\mathbf{v} &= \mathbf{u}\cdot\mathbf{u} - 2\alpha\mathbf{e}_1\cdot\mathbf{u} + \alpha^2\mathbf{e}_1\cdot\mathbf{e}_1 \\
&= 2\mathbf{u}\cdot\mathbf{u} - 2\alpha\mathbf{e}_1\cdot\mathbf{u} \\
&= 2(\mathbf{u} - \alpha\mathbf{e}_1)\cdot\mathbf{u} \\
&= 2\mathbf{v}\cdot\mathbf{u} \\
\mathbf{Q}\mathbf{u} &= \left(\mathbf{I} - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T\right)\mathbf{u} \\
&= \mathbf{u} - \frac{2\mathbf{v}^T\mathbf{u}}{\mathbf{v}^T\mathbf{v}}\mathbf{v} \\
&= \mathbf{u} - \frac{\mathbf{v}^T\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\mathbf{v} \\
&= \mathbf{u} - \mathbf{v} \\
&= \alpha\mathbf{e}_1
\end{aligned}
$$

  - Sign of α matters: $\mathbf{u} = \pm\mathbf{e}_1$ must not produce $\mathbf{v} = \mathbf{0}$ (!!)
  - Note the pattern of zeros

$$
\mathbf{Q}\begin{pmatrix} * \\ * \\ * \\ * \end{pmatrix} = \begin{pmatrix} * \\ 0 \\ 0 \\ 0 \end{pmatrix}
$$

  - We can also apply to a subset of a vector
    - Householder on a smaller vector, with identity on top

$$
\mathbf{Q}\begin{pmatrix} * \\ * \\ * \\ * \end{pmatrix} = \begin{pmatrix} * \\ * \\ 0 \\ 0 \end{pmatrix} \qquad \Longrightarrow \qquad \mathbf{Q} = \begin{pmatrix} 1 & \\ & \overline{\mathbf{Q}} \end{pmatrix}
$$

$$
\mathbf{Q}\begin{pmatrix} * \\ * \\ * \\ * \end{pmatrix} = \begin{pmatrix} * \\ * \\ * \\ 0 \end{pmatrix} \qquad \Longrightarrow \qquad \mathbf{Q} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \overline{\mathbf{Q}} \end{pmatrix}
$$

    - Leaves top entries untouched

- Algorithm (key: target entries, modified entries)

$$\mathbf{Q}_1\mathbf{A} = \mathbf{Q}_1 \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} = \begin{pmatrix} * & * & * & * \\ & * & * & * \\ & * & * & * \\ & * & * & * \end{pmatrix}$$

$$\mathbf{Q}_2\mathbf{Q}_1\mathbf{A} = \mathbf{Q}_2 \begin{pmatrix} * & * & * & * \\ & * & * & * \\ & * & * & * \\ & * & * & * \end{pmatrix} = \begin{pmatrix} * & * & * & * \\ & * & * & * \\ & & * & * \\ & & * & * \end{pmatrix}$$

$$\mathbf{Q}_3\mathbf{Q}_2\mathbf{Q}_1\mathbf{A} = \mathbf{Q}_3 \begin{pmatrix} * & * & * & * \\ & * & * & * \\ & & * & * \\ & & * & * \end{pmatrix} = \begin{pmatrix} * & * & * & * \\ & * & * & * \\ & & * & * \\ & & & * \end{pmatrix} = \mathbf{R}$$

$$\mathbf{Q}_3\mathbf{Q}_2\mathbf{Q}_1\mathbf{A} = \mathbf{R}$$

$$\mathbf{A} = \underbrace{\mathbf{Q}_1^T\mathbf{Q}_2^T\mathbf{Q}_3^T}_{\mathbf{Q}}\mathbf{R}$$

5 Frobenius Norm

- Definition

$$\|\mathbf{A}\|_F = \sqrt{\sum_{ij} A_{ij}^2}$$

$$\|\mathbf{A}\|_F^2 = \sum_{ij} A_{ij}^2$$

- Looks like L^2 norm for vectors
- Simple

$$\|c\mathbf{A}\|_F = |c| \ \|\mathbf{A}\|_F$$

$$\|\mathbf{A}\|_F \geq 0$$

$$\|\mathbf{A}\|_F = 0 \qquad \text{if and only if } \mathbf{A} = \mathbf{0}$$

- Triangle inequqality

$$\|\mathbf{A}+\mathbf{B}\|_F = \|\mathbf{A}\|_F + \|\mathbf{B}\|_F \qquad \text{Flatten the matrix into a vector}$$

- Equivalent expression

$$\begin{aligned}
\|\mathbf{A}\|_F^2 &= \mathrm{tr}(\mathbf{A}^T\mathbf{A}) \\
\mathrm{tr}(\mathbf{M}) &= \sum_i M_{ii} \\
(\mathbf{A}^T\mathbf{A})_{ij} &= \sum_k A_{ki}A_{kj} \\
\mathrm{tr}(\mathbf{A}^T\mathbf{A}) &= \sum_{ki} A_{ki}A_{ki} = \|\mathbf{A}\|_F^2 \\
\|\mathbf{A}\|_F^2 &= \sum_{ij} A_{ij}^2 \\
\|\mathbf{A}\|_F^2 &= \mathrm{tr}(\mathbf{A}\mathbf{A}^T) \qquad \text{Since } \mathrm{tr}(\mathbf{A}\mathbf{M}) = \mathrm{tr}(\mathbf{M}\mathbf{A})
\end{aligned}$$

- Propery of trace:

$$\begin{aligned}
\mathrm{tr}(\mathbf{A}\mathbf{M}) &= \mathrm{tr}(\mathbf{M}\mathbf{A}) \\
(\mathbf{A}\mathbf{M})_{ik} &= \sum_j A_{ij}M_{jk} \\
\mathrm{tr}(\mathbf{A}\mathbf{M}) &= \sum_{ij} A_{ij}M_{ji} \\
(\mathbf{M}\mathbf{A})_{ik} &= \sum_j M_{ij}A_{jk} \\
\mathrm{tr}(\mathbf{M}\mathbf{A}) &= \sum_{ij} M_{ij}A_{ji} \\
\|\mathbf{A}\|_F^2 &= \mathrm{tr}(\mathbf{A}^T\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{A}^T)
\end{aligned}$$

- Rotation invariance

$$\begin{aligned}
\|\mathbf{U}\mathbf{A}\|_F^2 &= \mathrm{tr}((\mathbf{U}\mathbf{A})^T(\mathbf{U}\mathbf{A})) \\
&= \mathrm{tr}(\mathbf{A}^T\mathbf{U}^T\mathbf{U}\mathbf{A}) \\
&= \mathrm{tr}(\mathbf{A}^T\mathbf{A}) \\
&= \|\mathbf{A}\|_F^2 \\
\|\mathbf{A}\mathbf{U}\|_F^2 &= \mathrm{tr}((\mathbf{A}\mathbf{U})(\mathbf{A}\mathbf{U})^T) \\
&= \mathrm{tr}(\mathbf{A}\mathbf{U}\mathbf{U}^T\mathbf{A}^T) \\
&= \mathrm{tr}(\mathbf{A}\mathbf{A}^T) \\
&= \|\mathbf{A}\|_F^2 \\
\|\mathbf{A}\|_F^2 &= \|\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\|_F^2 \\
&= \|\boldsymbol{\Sigma}\mathbf{V}^T\|_F^2 \\
&= \|\boldsymbol{\Sigma}\|_F^2 \\
&= \sum_i \sigma_i^2
\end{aligned}$$

- Remember Cauchy Schwarz

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

- Matrix-vector

$$\begin{aligned}
\|\mathbf{A}\mathbf{u}\| &\leq \|\mathbf{A}\|_F \|\mathbf{u}\| \\
\mathbf{A} &= \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \mathbf{v}_3^T \end{pmatrix} \\
\mathbf{A}\mathbf{u} &= \begin{pmatrix} \mathbf{v}_1^T \mathbf{u} \\ \mathbf{v}_2^T \mathbf{u} \\ \mathbf{v}_3^T \mathbf{u} \end{pmatrix} \\
\|\mathbf{A}\mathbf{u}\|^2 &= (\mathbf{v}_1 \cdot \mathbf{u})^2 + (\mathbf{v}_2 \cdot \mathbf{u})^2 + (\mathbf{v}_3 \cdot \mathbf{u})^2 \\
&\leq \|\mathbf{v}_1\|^2 \|\mathbf{u}\|^2 + \|\mathbf{v}_2\|^2 \|\mathbf{u}\|^2 + \|\mathbf{v}_3\|^2 \|\mathbf{u}\|^2 \\
&= (\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + \|\mathbf{v}_3\|^2) \|\mathbf{u}\|^2 \\
&= \|\mathbf{A}\|_F^2 \|\mathbf{u}\|^2
\end{aligned}$$

- Matrix-matrix

$$\begin{aligned}
\|\mathbf{A}\mathbf{B}\|_F &\leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F \\
\mathbf{B} &= \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} \\
\mathbf{A}\mathbf{B} &= \begin{pmatrix} \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \mathbf{A}\mathbf{v}_3 \end{pmatrix} \\
\|\mathbf{A}\mathbf{B}\|_F^2 &= \|\mathbf{A}\mathbf{v}_1\|^2 + \|\mathbf{A}\mathbf{v}_2\|^2 + \|\mathbf{A}\mathbf{v}_3\|^2 \\
&\leq \|\mathbf{A}\|_F^2 \|\mathbf{v}_1\|^2 + \|\mathbf{A}\|_F^2 \|\mathbf{v}_2\|^2 + \|\mathbf{A}\|_F \|\mathbf{v}_3\|^2 \\
&= \|\mathbf{A}\|_F^2 (\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + \|\mathbf{v}_3\|^2) \\
&= \|\mathbf{A}\|_F^2 \|\mathbf{B}\|_F^2
\end{aligned}$$