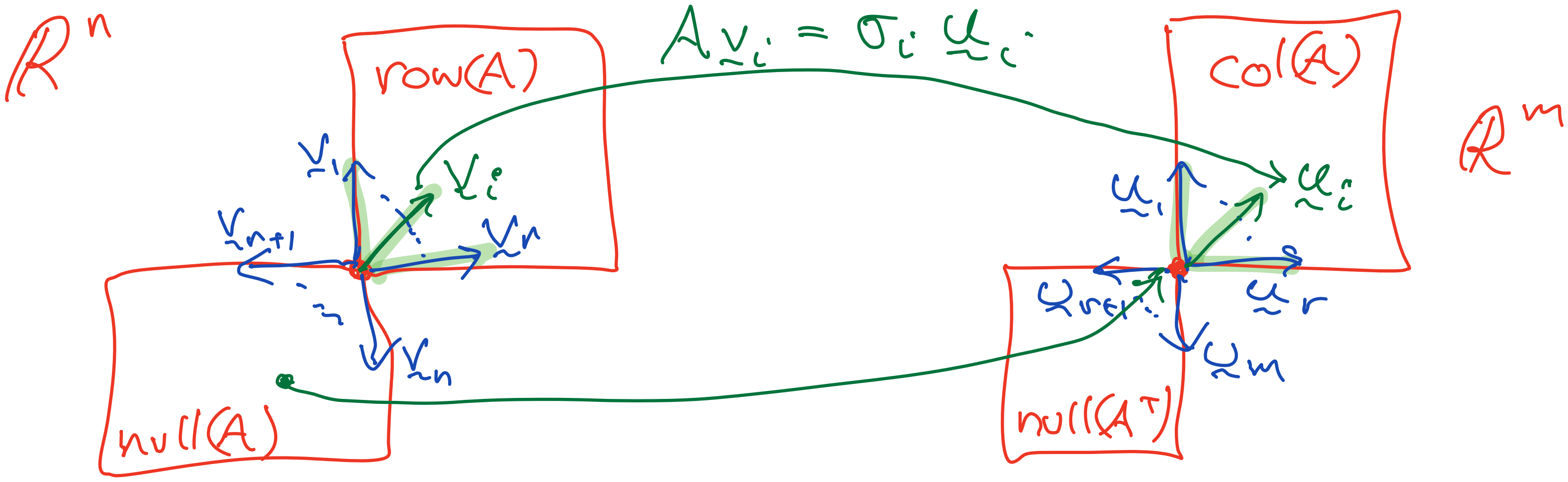
Theorem 8.3.8 (Existence of SVD). *Every matrix has a singular value decomposition.* proof (MTH 2025)

Theorem 8.3.9. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $T(\mathbf{x}) = A\mathbf{x}$, and let A have singular value decomposition $A = U\Sigma V^T$. If B is the set of columns of V and B' is the set of columns of U, then $[T]_{B',B} = \Sigma$.* $T : R^n \to R^m$



proof: Let $B = \{v_1, \dots, v_n\}$ & $B' = \{u_1, \dots, u_m\}$

Then $[T]_{B',B} = \left[[T(v_1)]_{B'} \cdots [T(v_n)]_{B'} \right]$

But $T(v_i) = Av_i = U(\underline{\Sigma V^T v_i})$

$= U c_i$ where $c_i = \Sigma V^T v_i \in R^m$

$= \sum_{j=1}^{n} (c_i)_j u_j$

$\Rightarrow$ $[T(v_i)]_{B'} = c_i = \Sigma V^T v_i = \Sigma (v_i^T V)^T$

But $(v_i^T V)_j = \sum_{k=1}^{n} (v_i)_k (V)_{kj} = \sum_{k=1}^{n} (v_i)_k (v_j)_k = v_i \cdot v_j$

$= \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases} = (e_i^T)_j$

$\Rightarrow$ $v_i^T V = e_i^T$ $\Rightarrow$ $[T(v_i)]_{B'} = \Sigma e_i \quad \forall i$

$\Rightarrow \quad [T]_{B'B} = I$ $\square$

Theorems 8.3.6 and 8.3.8 justify the following algorithm.

Algorithm 8.3.10 (SVD)**.** *For an $m \times n$ matrix A:*

(i) *Find the eigenvalues λ_i and corresponding orthonormal set of eigenvectors $\mathbf{v}_i$ of $A^T A$. Order them so that $\lambda_1 \geq \ldots \geq \lambda_r > 0$ and $\lambda_i = 0$ for $i > r$.*

(ii) *Let $V = \begin{bmatrix} \mathbf{v}_1 \ldots \mathbf{v}_n \end{bmatrix}$*

(iii) *Let the nonzero diagonal entries of Σ be $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \ldots, r$*

$\Leftrightarrow A \underset{\sim}{v}_i = \sigma_i \underset{\sim}{u}_i$

(iv) *Let $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\sigma_i}$ for $i = 1, \ldots, r$*

(v) *Let $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ be an orthonormal basis for $\mathrm{null}(A^T)$ and set $U = \begin{bmatrix} \mathbf{u}_1 \ldots \mathbf{u}_m \end{bmatrix}$*

~~**Definition 8.3.11.** *An eigenvector of $A^T A$ is a* ***right singular vector*** *of A. An eigenvector of AA^T is a* ***left singular vector*** *of*~~ *A.*

~~**Lemma 8.3.12.** *Let A have SVD $A = U\Sigma V^T$. The columns*~~ *of* ~~*U are left singular vectors of A, and the columns of V are right singular vectors of A.*~~

Exercise 8.3.13. *Find a singular value decomposition of*

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$A^TA = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\det(\lambda I - A^TA) = \lambda(\lambda-1)(\lambda-2)$

so eigenvalues of A^TA are $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 0$

$\underline{\underline{\lambda_1}}$ $\text{null}(2I - A^TA) = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}$

$\underline{\underline{\lambda_2}}$ $\text{null}(I - A^TA) = \text{span}\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$

$\underline{\underline{\lambda_3}}$ $\text{null}(0I - A^TA) = \text{span}\left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\right\}$

The eigenspaces are 1-dimensional (so no GS)

$\Rightarrow$ we need only normalise. Let

$\underset{\sim}{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\underset{\sim}{v}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, $\underset{\sim}{v}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

& $V = [\underset{\sim}{v}_1 \ \underset{\sim}{v}_2 \ \underset{\sim}{v}_3] = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$

& $\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

$\underset{\sim}{u}_1 = \frac{1}{\sigma_1} A \underset{\sim}{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

so $U = [\underset{\sim}{u}_1 \ \underset{\sim}{u}_2] = I$

$\underset{\sim}{u}_2 = \frac{1}{\sigma_2} A \underset{\sim}{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$\Rightarrow$ A has SVD $A = U \Sigma V^T$.

OPTIONAL - NOT ASSESSED

There is an alternative form of the SVD, analogous to the spectral decomposition.

Theorem 8.3.14. *Let A be an $m \times n$ matrix of rank $r > 0$ with SVD $A = U\Sigma V^T$, where $U = \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_m\end{bmatrix}$, $V = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_n\end{bmatrix}$ and $(\Sigma)_{ii} = \sigma_i$ for $1 \leq i \leq r$. Then*

$$A = \sigma_1\,\mathbf{u}_1\,\mathbf{v}_1^T + \ldots + \sigma_r\,\mathbf{u}_r\,\mathbf{v}_r^T. \tag{8.1}$$

This can be written in matrix form $A = \tilde{U}\,\tilde{\Sigma}\,\tilde{V}^T$ where $\tilde{U} = \begin{bmatrix}\mathbf{u}_1 \ldots \mathbf{u}_r\end{bmatrix}$, $\tilde{V} = \begin{bmatrix}\mathbf{v}_1 \ldots \mathbf{v}_r\end{bmatrix}$ and $\tilde{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$.

Definition 8.3.15. *We call* (8.1) *the* ***singular value expansion****, and its matrix form the* ***reduced singular value decomposition****.*

Proof of Theorem 8.3.14. Let $1 \leq i \leq m$, $1 \leq j \leq n$

Then

$$(A)_{ij} = (U\Sigma V^T)_{ij} = \sum_{k=1}^{m}\sum_{l=1}^{n} (U)_{ik}(\Sigma)_{kl}(V)_{jl}$$

$$= \sum_{k=1}^{r}\sum_{l=1}^{r} (U)_{ik}(\Sigma)_{kl}(V)_{jl} \quad \text{since } (\Sigma)_{kl} = 0 \ \forall\, k, l > r$$

$$= \sum_{k=1}^{r}\sum_{l=1}^{r} (\tilde{U})_{ik}(\tilde{\Sigma})_{kl}(\tilde{V})_{jl}$$

$$= (\tilde{U}\tilde{\Sigma}\tilde{V}^T)_{ij} \Rightarrow A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$$

Moreover, $(A)_{ij} = \sum_{k=1}^{r} (\tilde{U})_{ik}\,\sigma_k\,(\tilde{V})_{jk}$

$$= \sum_{k=1}^{r} (\underset{\sim}{u}_k)_i\,\sigma_k\,(\underset{\sim}{v}_k)_j$$

$$= \sum_{k=1}^{r} \sigma_k \, (\underset{\sim}{u}_k)_i \, (\underset{\sim}{v}_k^T)_j$$

$$= \sum_{k=1}^{r} \sigma_k \, (\underset{\sim}{u}_k \underset{\sim}{v}_k^T)_{ij}$$

$$= \left(\sum_{k=1}^{r} \sigma_k \, \underset{\sim}{u}_k \underset{\sim}{v}_k^T \right)_{ij}$$

$$\Rightarrow A = \sum_{k=1}^{r} \sigma_k \, \underset{\sim}{u}_k \underset{\sim}{v}_k^T$$

$\square$

$\square$

8.4 Consequences of SVD

8.4.1 Condition number

OMITTED

Definition 8.4.1. *The* ***condition number*** *of an invertible $n \times n$ matrix is the ratio between its largest and smallest singular values*

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}.$$

If $\kappa(A)$ is very large, we say A is **ill-conditioned**. In practice, this occurs when $\kappa(A)$ is larger than the reciprocal of the machine's precision. It is much harder to numerically solve a linear system $A\mathbf{x} = \mathbf{b}$ when A is ill-conditioned.

E.g. 8.4.2. *Find $\kappa(A)$ for*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \epsilon \end{bmatrix}.$$

For what regime of ϵ is A ill-conditioned?

8.4.2 Pseudoinverse

Definition 8.4.3. *Let Σ be a diagonal $m \times n$ matrix with precisely r nonzero entries: $(\Sigma)_{ii} = \sigma_i > 0$ for $1 \leq i \leq r$. The* ***pseudoinverse*** *of Σ, denoted Σ^+, is the diagonal $n \times m$ matrix with nonzero entries $(\Sigma^+)_{ii} = 1/\sigma_i$ for $1 \leq i \leq r$. If A is any $m \times n$ matrix with SVD $A = U\Sigma V^T$, we define its* ***pseudoinverse*** *to be $A^+ = V\Sigma^+U^T$.*

Lemma 8.4.4. *Every matrix has a unique pseudoinverse.*

Proof. ∎

OMITTED

Lemma 8.4.5. *If A is an $m \times n$ matrix with reduced SVD $A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$, then $A^+ = \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^T$.*

Proof. ∎

Lemma 8.4.6. *For any matrix A we have* $\operatorname{rank}(A) = \operatorname{rank}(A^+)$.

Proof. Prob. Set. ∎

Lemma 8.4.7. *If A is invertible, then $A^+ = A^{-1}$.*

Proof. Prob. Set ∎

E.g. 8.4.8. *If* $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$, *find* A^+.

$A^TA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has eigenvalues 3 & 1 & eigenspace bases $\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$ & $\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}$.

$\Rightarrow V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$, $\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$

$\underset{\sim}{u}_1 = \frac{1}{\sigma_1} A \underset{\sim}{v}_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$, $\underset{\sim}{u}_2 = \frac{1}{\sigma_2} A \underset{\sim}{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$

Find $\underset{\sim}{u}_3$ by finding a basis for $null(A^T)$

$U = [\underset{\sim}{u}_1, \underset{\sim}{u}_2, \underset{\sim}{u}_3]$

$A^+ = V \Sigma^+ U^T = V \begin{pmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} U^T$

Theorem 8.4.9. *If A is an $m \times n$ matrix then:*

(a) $\operatorname{col}(A^+) = \operatorname{row}(A)$

(b) $\operatorname{row}(A^+) = \operatorname{col}(A)$ follows from (a) because $A^{++} = A$

(c) $\ker(A^+) = \operatorname{coker}(A)$

(d) $\operatorname{coker}(A^+) = \ker(A)$ " " (c) " "

Proof. ✱ OPTIONAL - NOT ASSESSED ✱

Let A have RSVD $A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$.

(a) If $\underset{\sim}{x} \in \operatorname{col}(A^+)$, then

$\underset{\sim}{x} = A^+ \underset{\sim}{b}$ for some $\underset{\sim}{b} \in \mathbb{R}^m$.

But $A^+\underset{\sim}{b} = \tilde{V}(\tilde{\Sigma}^{-1}\tilde{U}^T \underset{\sim}{b})$

$= \tilde{V}\underset{\sim}{c}$ where $\underset{\sim}{c} = \tilde{\Sigma}^{-1}\tilde{U}^T\underset{\sim}{b} \in \mathbb{R}^r$

$\Rightarrow A^+\underset{\sim}{b} = \sum_{i=1}^{r} c_i \underset{\sim}{v}_i \in \operatorname{row}(A)$ where $r = \operatorname{rank}(A)$

since $\{\underset{\sim}{v}_1, \dots, \underset{\sim}{v}_r\}$ is a basis for $\operatorname{row}(A)$

Therefore $\operatorname{col}(A^+) \subseteq \operatorname{row}(A)$.

But Lemma 8.4.6 $\Rightarrow \operatorname{rank}(A^+) = \operatorname{rank}(A)$

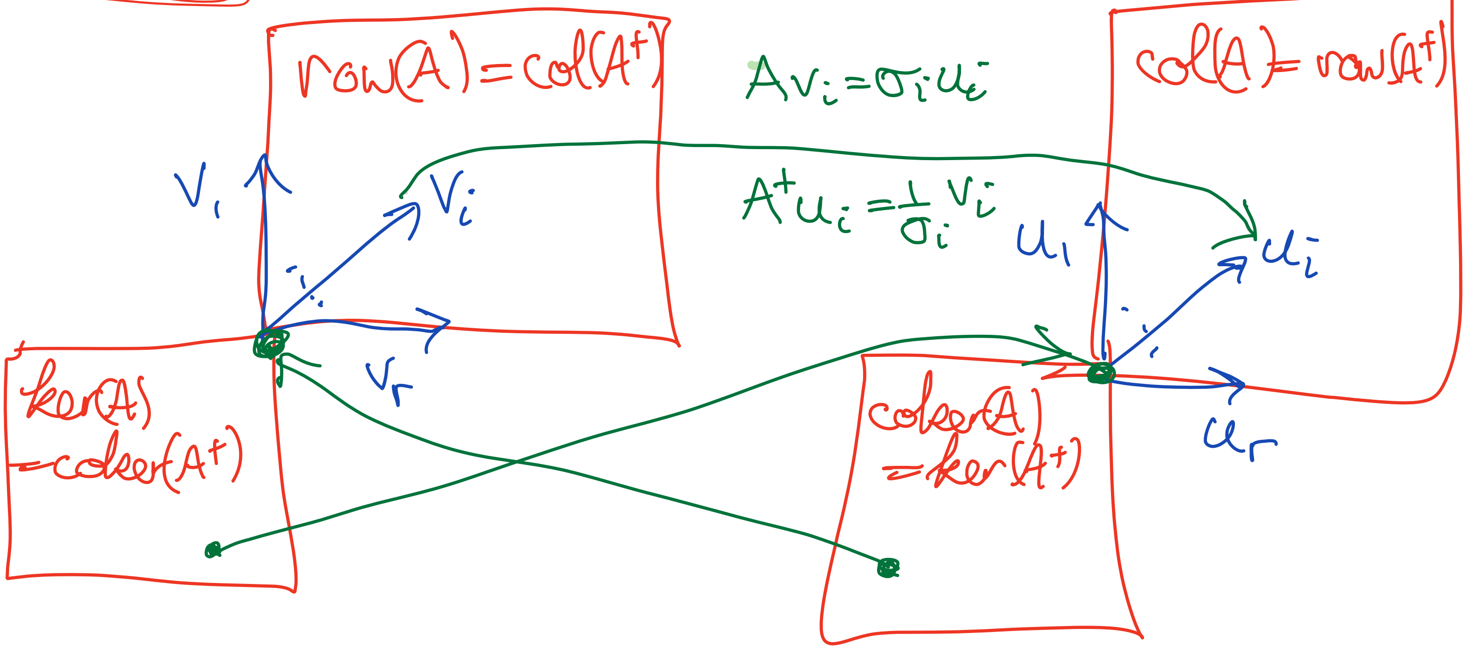
$\Rightarrow \dim \operatorname{col}(A^+) = \dim \operatorname{row}(A)$

so $\operatorname{col}(A^+) = \operatorname{row}(A)$ by Th$^{\text{m}}$ 4.3.17

(c) $\ker(A^+) = \operatorname{row}(A^+)^\perp = \operatorname{col}(A)^\perp = \operatorname{coker}(A)$ $\square$

Theorem 8.4.10. *Let A be an $m \times n$ matrix, and define $T_A : \text{row}(A) \to \text{col}(A)$ and $T_{A^+} : \text{row}(A^+) \to \text{col}(A^+)$ via $T_A(\mathbf{x}) = A\mathbf{x}$ and $\boxed{T_{A^+}(\mathbf{x}) = A^+\mathbf{x}}$. Then T_A is an isomorphism, and $\boxed{T_A^{-1} = T_{A^+}}$.*

Proof: Prob Set



(Show on prob set that $A^+ \underset{\sim}{u}_i = \frac{1}{\sigma_i} \underset{\sim}{v}_i$)

(*OPTIONAL*)

Theorem 8.4.11. *The system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution $\mathbf{x}_*$ of minimal length given by*

$$\mathbf{x}_* = A^+\mathbf{b}.$$

Proof. It can be verified directly (Prob Set) that $\underset{\sim}{x}_*$ satisfies normal eq$^{\underline{n}}$s for A. So Th$^{\underline{m}}$ 1.2.13 $\Rightarrow$ every sol$^{\underline{n}}$ of the normal eq$^{\underline{n}}$s can be written as $\underset{\sim}{x}_* + \underset{\sim}{z}$ for some $\underset{\sim}{z} \in \text{null}(A^TA)$. But $\text{null}(A^TA) = \text{null}(A)$ $\Rightarrow \underset{\sim}{z} \in \text{null}(A)$. Moreover,

$$A^+\underset{\sim}{b} \in \text{col}(A^+) = \text{row}(A) \text{ by Th}^{\underline{m}}\ 8.4.9$$

$$\Rightarrow \underset{\sim}{x}_* \in \text{row}(A)$$

Since $\text{null}(A) \perp \text{row}(A)$, Pythagoras' th$^{\underline{m}}$ $\Rightarrow$

$$\|\underset{\sim}{x}_* + \underset{\sim}{z}\|^2 = \|\underset{\sim}{x}_*\|^2 + \|\underset{\sim}{z}\|^2 > \|\underset{\sim}{x}_*\|^2 \quad \forall\ \underset{\sim}{z} \neq \underset{\sim}{0}$$

So any other LS sol$^{\underline{n}}$ has larger length. $\square$

(*OPTIONAL*)

E.g. 8.4.12. *Find the minimum length least squares solution of* $A\mathbf{x} = \mathbf{b}$ *where*

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Find SVD for A

$$\Rightarrow A^+ = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

so $$x_* = A^+ \underset{\sim}{b} = \frac{1}{4}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

8.5 Applications of SVD (A&R §9.5)

NOT ASSESSED

8.5.1 Data compression (A&R §9.5)

Digital images are stored as matrices. Each matrix entry corresponds to a given pixel. Each possible pixel colour/shade corresponds to a fixed number. For example, greyscale images typically use 256 pixel values ranging from 0 (white) to 255 (black).

Consider a digital image encoded by an $m \times n$ matrix A. We would like to store the matrix using as little memory as possible. Suppose each entry requires b bytes of storage. Naively storing each entry requires $m\,n\,b$ bytes.

If A has a singular value expansion

$$A = \sigma_1\,\mathbf{u}_1\,\mathbf{v}_1^T + \ldots + \sigma_r\,\mathbf{u}_r\,\mathbf{v}_r^T\,,$$

then we can instead store the singular values σ_i and vectors $\mathbf{u}_i$, $\mathbf{v}_i$ for $1 \leq i \leq r$. This requires storing $r(1 + m + n)$ entries, i.e. $r(1 + m + n)b$ bytes. If $r \ll \min(m, n)$ this may give a significant saving.

But if $\sigma_{k+1}, \ldots \sigma_r$ are all *small*, we may achieve a satisfactory approximation to the matrix A by retaining only the first k terms. Storing this approximation requires only $k(1+m+n)b$ bytes. Such an approximation gives a simple technique for data compression of an image.

NOT ASSESSED

E.g. storing a 50 term approximation of a 1000 by 1000 matrix requires storing 10^5 entries instead of 10^6; implying the image is compressed by a factor of 10.

(a) 4 terms (b) 20 terms (c) 50 terms

8.5.2 Principle Component Analysis

Suppose we have n experimental measurements of p variables, y_{ij} with $1 \leq i \leq n$, $1 \leq j \leq p$. Define the $n \times p$ matrix X by

$$(X)_{ij} = y_{ij} - \frac{1}{n} \sum_{i=1}^{n} y_{ij}.$$

The singular value decomposition of the data matrix X is one of the most widely used tools in descriptive statistics.

The **sample covariance matrix** of the data is $S = (n-1)^{-1} X^T X$.

The ith **principle component** is the eigenvector of S corresponding to the ith largest eigenvalue.

The ith eigenvalue of S is the **variance** of the ith principle component.

NOT ASSESSED

From a statistical ~~perspective, the orthogonality~~ of the principle components implies they are uncorrelated.

If p is large, understanding the correlations between all the variables is a difficult task. Principle component analysis involves focusing only on those principle components with large variance, and discarding those with small variance. This can yield a dramatic dimension reduction of the data, albeit an approximate one.