Section 6.7

Let A be an $m \times n$ matrix, so that multiplication by it is a function from $\mathbb{R}^n$ into $\mathbb{R}^m$, and multiplication by A^T is a function from $\mathbb{R}^m$ int $\mathbb{R}$.. Recall that

$$\begin{aligned} A\mathbf{x} = \mathbf{0}_m &\implies A^T A\mathbf{x} = \mathbf{0}_n \\ &\implies ||A\mathbf{x}||^2 = \mathbf{x}^T A^T A\mathbf{x} = 0 \\ &\implies A\mathbf{x} = \mathbf{0}_m \end{aligned}$$

so A and $A^T A$ have the same nullspace. Similarly, so do A^T and AA^T. Also, because $\text{rank}(A) = \text{rank}(A^T)$, we get

$$r = \text{rank}(A^T A) = \text{rank}(A) = \text{rank}(A^T) = \text{rank}(AA^T) \ .$$

$A^T A$ and AA^T are square, symmetric, positive semidefinite matrices, so they have orthonormal diagonalizations. We start with $A^T A$ and eventually drag in AA^T:

$A^T A$ is $n \times n$, so there is an $n \times n$ orthogonal matrix V with columns the orthonormal eigvecs of $A^T A$, and an $n \times n$ diagonal matrix D with nonnegative diagonal entries

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2 \ ,$$

the eigvals of $A^T A$ (and $\sigma_i \geq 0$).

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the columns of V. Temporarily, let r be the subscript of the last nonzero σ_i. Then $A^T A\mathbf{v}_1, \ldots, A^T A\mathbf{v}_r$ are just nonzero multiples of $\mathbf{v}_1, \ldots, \mathbf{v}_r$; so the first r $\mathbf{v}$'s are in the row space of A. And $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ are in the nullspace of A. But together they form a basis for $\mathbb{R}^n$; so the first r $\mathbf{v}$'s are a basis for the row space of A, and $r = \text{rank}(A)$, while the rest of the $\mathbf{v}$'s form a basis for the nullspace of A. The first r σ's (the nonzero ones) are the **singular values** of A.

Now we bring back AA^T: For each $i \leq r$,

$$(AA^T)(A\mathbf{v}_i) = A(A^TA)\mathbf{v}_i = A(\sigma_i^2\mathbf{v}_i) = \sigma_i^2(A\mathbf{v}_i) \ ,$$

so σ_i^2 is an eigval of AA^T, too, with corresponding eigvec $A\mathbf{v}_i$.

Also, if $i \neq j$, we have $\mathbf{v}_i \cdot \mathbf{v}_j = 0$, so

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \mathbf{v}_i^T(A^TA\mathbf{v}_j) = \mathbf{v}_i^T(\sigma_j^2\mathbf{v}_j) = \sigma_j^2(\mathbf{v}_i \cdot \mathbf{v}_j) = 0 \ .$$

So $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ are r orthogonal, nonzero (and hence lin ind) vectors in col space of A, so they form a basis. If we divide them by their lengths and throw in an orthonormal basis for the left nullspace of A, we'll have an orthonormal basis for $\mathbb{R}^n$.

The $\mathbf{v}$'s are unit vectors, so

$$||A\mathbf{v}_i|| = \sqrt{(A\mathbf{v}_i) \cdot (A\mathbf{v}_i)} = \sqrt{\mathbf{v}_i^T A^T A\mathbf{v}_i} = \sqrt{\sigma_i^2(\mathbf{v}_i \cdot \mathbf{v}_i)} = \sigma_i \ .$$

So to get an orthonormal basis of $\mathbb{R}^m$ consisting of eigvecs of AA^T, we set $\mathbf{u}_i = A\mathbf{v}_i/\sigma_i$ for $i = 1, \ldots, r$, and throw in an orthonormal basis $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ of the left nullspace of A (which are eigvecs of AA^T with eigval 0). Use these $\mathbf{u}$'s as the columns of an orthogonal matrix U; then $AA^T = U\widehat{D}U^T$, where $\widehat{D}$ is an $m \times m$ diagonal matrix with diagonal entries $\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0$.

Singular Value Decomposition

Let Σ denote the $m \times n$ matrix with $\sigma_1, \ldots, \sigma_r$ down the start of its main diagonal and 0's everywhere else. Then $D = \Sigma^T \Sigma$ and $\widehat{D} = \Sigma \Sigma^T$.

Theorem

$A = U \Sigma V^T$.

Proof.

We only need to check that these two matrices multiply all the elements of some basis of $\mathbb{R}^n$ to the same things in $\mathbb{R}^m$, so we can check it on the $\mathbf{v}$'s:
$V^T \mathbf{v}_i = V^{-1} \mathbf{v}_i$ is the i-th col of the identity, and Σ multiplies that by σ_i if $i \leq r$ and by zero if $i > r$. Then U multiplies that to $\sigma_i \mathbf{u}_i = \sigma_i (A\mathbf{v}_i / \sigma_i) = A\mathbf{v}_i$ if $i \leq r$ and to the zero vector if $i > r$; but that is again $A\mathbf{v}_i$. $\square$

Algebraic consequences of SVD

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T\Sigma V^T = VDV^T \quad \text{(Check!)}$$
$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma\Sigma^T U^T = U\widehat{D}U^T \quad \text{(Check!)}$$
$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \text{ for } i \leq r \quad \text{(by def of } \mathbf{u}_i\text{)}$$
$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

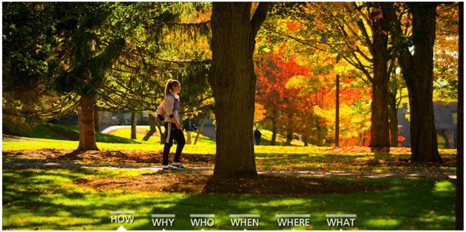
Each $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is a rank-1 matrix, so it can be rebuilt from $m+n+1$ numbers (first row, first column, and σ_i), vs. mn for A. For big values of m, n, even if we only use the first, say, 80 terms of the sum, $80(m+n+1)$ is still much smaller than mn, so less storage is required. And the sum of the first 80 terms is "the best rank-80 approximation to A."

The text answers the question:
When is the singular value decomposition the same as the diagonalization?

- To have a diagonalization, A must be square, so all matrices are square of the same size.
- Suppose $A = U\Sigma V^T$, where $V^T = U^{-1} = U^T$. Then A is symmetric.
- And diagonal entries in Σ are nonnegative, so A is positive semidefinite.

Application: .png image compression

Original photo:



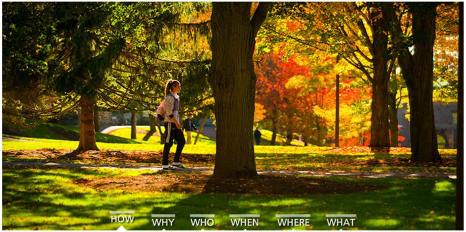
527 KB

708×352 pixels, each at least 3 memory cells (different color levels), so with no compression, 730 KB of memory.

If it is stored as a $708 \times ((352)(3) = 1056)$ matrix, it could have rank up to 708, i.e., up to 708 singular values. Suppose we use s singular values to approximate the matrix. Then we need to store $M(s) = s(708 + 1056 + 1) = 1765s$ numbers.

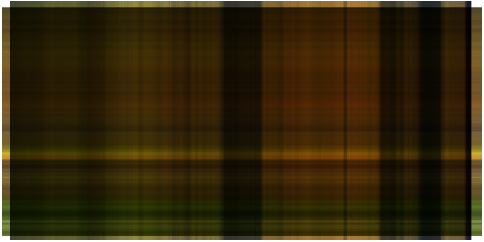
Application: .png image compression

Original photo:



527 KB

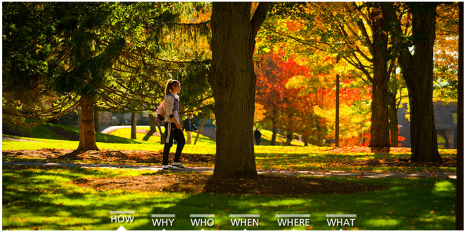
Number of singular values used: 1



161 KB, $M(1) \approx 1.7$ KB

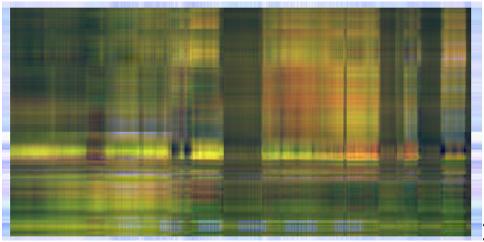
Application: .png image compression

Original photo:



527 KB

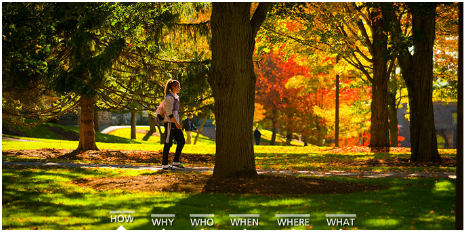
Number of singular values used: 5



263 KB, $M(5) \approx 8.6$ KB

Application: .png image compression

Original photo:



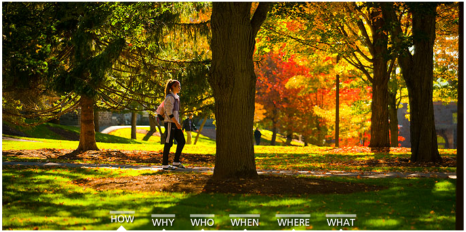
527 KB

Number of singular values used: 20

343 KB, $M(20) \approx 34.5$ KB

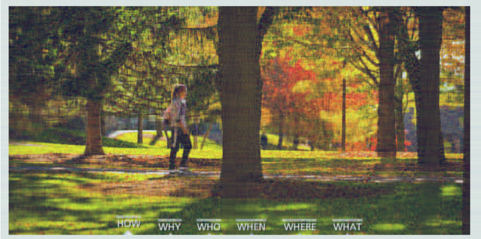
Application: .png image compression

Original photo:



527 KB

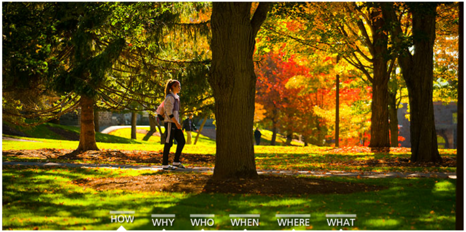
Number of singular values used: 50



410 KB, $M(50) \approx 86$ KB

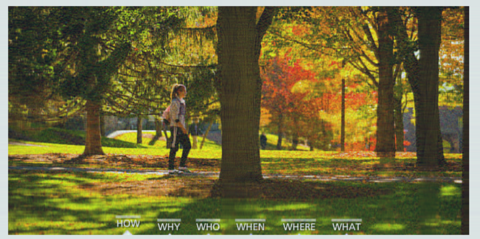
Application: .png image compression

Original photo:



527 KB

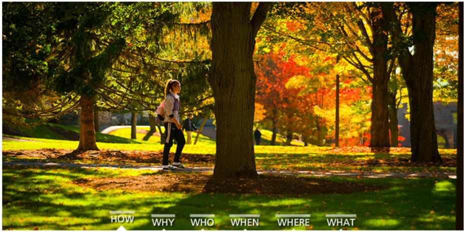
Number of singular values used: 80



446 KB, $M(80) \approx 138$ KB

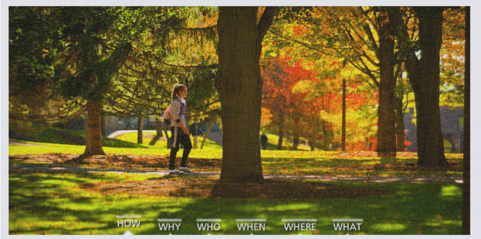
Application: .png image compression

Original photo:



527 KB

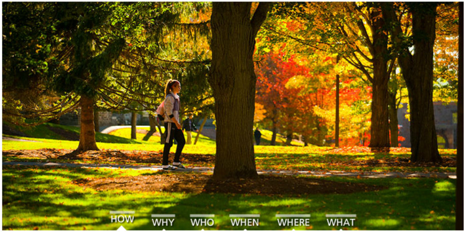
Number of singular values used: 100



462 KB, $M(100) \approx 172$ KB

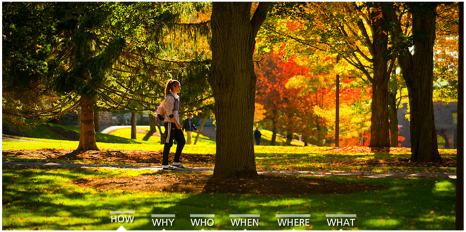
Application: .png image compression

Original photo:



527 KB

Number of singular values used: $1000 (> 708?)$



493 KB,

$M(1000) \approx 1724$ KB

Example (using R, but it's not needed)

```
> 15*A
      [,1]  [,2]
[1,]    10    30
[2,]    −7    24
[3,]   −34   −12
> eigen(t(A)%*%A)
$values
[1]  9  4
$vectors
     [,1]  [,2]
[1,]  0.6  -0.8
[2,]  0.8   0.6
> V←eigen(t(A)%*%A)$vectors; v1←V[,1]; v2←V[,2]
> u1←(1/3)*A%*%v1; u2←(1/2)*A%*%v2
```

Example, continued

```
> B←t(A); B
          [,1]       [,2]      [,3]
[1,] 0.6666667 −0.4666667 −2.266667
[2,] 2.0000000  1.6000000 −0.800000
> B[1,] ←B[1,]*3/2; B[2,] ← B[2,] − 2*B[1,]; B
     [,1] [,2] [,3]
[1,]    1 -0.7 -3.4
[2,]    0  3.0  6.0
> B[2,] ←B[2,]/3; B[1,] ←B[1,] + .7*B[2,]; B
     [,1] [,2] [,3]
[1,]    1    0   -2
[2,]    0    1    2
> u3←c(2,-2,1); u3←u3/sqrt(sum(t(u3)%*%u3)); u3
[1] 0.6666667 −0.6666667 0.3333333
```

Example, continued

```
> Sigma←cbind(c(3,0,0),c(0,2,0)); Sigma
     [,1] [,2]
[1,]    3    0
[2,]    0    2
[3,]    0    0
> U←cbind(u1,u2,u3); U%*%Sigma%*%t(V)
           [,1] [,2]
[1,]  0.6666667  2.0
[2,] −0.4666667  1.6
[3,] −2.2666667 −0.8
> A
           [,1] [,2]
[1,]  0.6666667  2.0
[2,] −0.4666667  1.6
[3,] −2.2666667 −0.8
```

Example, continued

```
> u1%*%t(v1)
     [,1]       [,2]
[1,]  0.4  0.5333333
[2,]  0.2  0.2666667
[3,] -0.4 -0.5333333
> u2%*%t(v2)
           [,1] [,2]
[1,] -0.2666667  0.2
[2,] -0.5333333  0.4
[3,] -0.5333333  0.4
> 3*u1%*%t(v1)+2*u2%*%t(v2)
           [,1] [,2]
[1,]  0.6666667  2.0
[2,] -0.4666667  1.6
[3,] -2.2666667 -0.8
```

Example by hand

Find the singular value decomposition $A = U\Sigma V$ for

$$A = \begin{bmatrix} 0 & 12/5 \\ 1 & 0 \\ 0 & 16/5 \end{bmatrix} .$$

Begin by diagonalizing

$$A^T A = \begin{bmatrix} 0 & 1 & 0 \\ 12/5 & 0 & 16/5 \end{bmatrix} \begin{bmatrix} 0 & 12/5 \\ 1 & 0 \\ 0 & 16/5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T :$$

Already diagonal, but we want the eigenvalues in decreasing order (and the eigenvectors, in same order, with length 1, but that's done here):

$$\sigma_1 = \sqrt{16} = 4, \ \mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \sigma_2 = \sqrt{1} = 1, \ \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} .$$

Now the first two $\mathbf{u}$'s:

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1 = \begin{bmatrix} 12/5 \\ 0 \\ 16/5 \end{bmatrix} /4 = \begin{bmatrix} 3/5 \\ 0 \\ 4/5 \end{bmatrix},$$

$$\mathbf{u}_2 = A\mathbf{v}_2/\sigma_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} /1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

The missing piece is an orthonormal basis for the left nullspace. But $\dim \mathbf{N}(A^T) = 3 - 2 = 1$, so all we need is either of the two length-1 vectors perpendicular to both $\mathbf{u}_1, \mathbf{u}_2$:

$$\begin{bmatrix} 3/5 & 0 & 4/5 \\ 0 & 1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 4/3 \\ 0 & 1 & 0 \end{bmatrix} \quad : \quad \mathbf{u}_3 = \begin{bmatrix} 4/5 \\ 0 \\ -3/5 \end{bmatrix}$$

So

$$U = \begin{bmatrix} 3/5 & 0 & 4/5 \\ 0 & 1 & 0 \\ 4/5 & 0 & -3/5 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \ V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} :$$

$$A = U\Sigma V^T \ .$$