Problem 1: Image compression and remembering John Nash

Figure 1: John Forbes Nash, Jr. (1928 – 2015). The image to be compressed.

If we were to rank the topics in computational linear algebra in terms of how frequently they arise in applications, singular value decomposition (SVD) would probably take one of the very top spots. In this problem, we introduce you to this concept and present one of its applications in image processing. The SVD also has applications in many other domains, notably in statistics (principal component analysis).

Let $A$ be a real $m \times n$ matrix of rank $r$. (Recall that the rank of $A$ is the number of linearly independent columns of $A$.) The singular value decomposition of $A$ is a decomposition of the form

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

where $U$, $\Sigma$, and $V$ are respectively $m \times m$, $m \times n$, and $n \times n$; $U$ and $V$ are orthogonal matrices (i.e., satisfy $U^T U = I$ and $V^T V = I$), and $\Sigma$ is a matrix with $r$ positive scalars $\sigma_1, \ldots, \sigma_r$ on the diagonal of its upper left $r \times r$ block and zeros everywhere else. The scalars $\sigma_1, \ldots, \sigma_r$ are called the singular values of $A$. They are given as

$$\sigma_i = \sqrt{i\text{-th eigenvalue of } A^T A},$$

and by convention they appear in descending order:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r.$$
The columns of $U$ and $V$ are respectively called the left and right \textit{singular vectors} of $A$ and can be obtained by taking an orthonormal set of eigenvectors\footnote{By an orthonormal set of vectors we mean a collection of vectors that are pairwise orthogonal and each have 2-norm equal to one.} for the matrices $AA^T$ and $A^TA$. In Matlab, the command \texttt{svd} handles these eigenvector computations for you and outputs the three matrices $U, \Sigma, V$.

First, you need to answer some very basic questions about the SVD.

1. (a) Show that eigenvalues of $A^TA$ are always nonnegative. (Hence singular values are well-defined as real, nonnegative scalars.)
   (b) Show that if $A$ is symmetric then the singular values of $A$ are the same as the absolute value of the eigenvalues of $A$.
   (c) Show that if $u_i$ and $u_j$ are eigenvectors of $A^TA$ associated to distinct eigenvalues $\lambda_i, \lambda_j$, then $u_i$ and $u_j$ are orthogonal.

\textbf{What does this have to do with optimization?} Let $A$ be a real $m \times n$ matrix with an SVD given by $A = U\Sigma V^T$ defined as above. For a positive integer $k \leq \min\{m, n\}$, we let $A_{(k)}$ denote an $m \times n$ matrix which is an “approximation” of the matrix $A$ obtained from its top $k$ singular values and singular vectors. Formally, we have

$$A_{(k)} = U_{(k)}\Sigma_{(k)}V_{(k)}^T,$$

where $U_{(k)}$ has the first $k$ columns of $U$, $V_{(k)}$ has the first $k$ columns of $V$, and $\Sigma_{(k)}$ is the upper left $k \times k$ block of $\Sigma$.

2. Show that

$$A_{(k)} = \min_{B \in \mathbb{R}^{m \times n}, \text{rank}(B) \leq k} \| A - B \|_2.$$

Here, $\| \cdot \|_2$ denotes the spectral norm of a matrix defined as $\| C \|_2 = \max_{\| x \|_2 = 1} \| Cx \|_2$. (Hint: You may want to first prove that the spectral norm of a matrix is \textit{unitarily invariant}, i.e., it does not change when the matrix is multiplied from left or right by an orthogonal matrix. You may also want to use the following fact from linear algebra: For any matrix $E \in \mathbb{R}^{m \times n}$, $\text{rank}(E) + \text{dim} \text{ null}(E) = n$.)

In words, the result you are proving states that among all $m \times n$ matrices of rank at most $k$, the matrix $A_{(k)}$ obtained from truncating the SVD is the one that best approximates $A$. 
in the spectral norm. The benefit in approximating a matrix with low-rank matrices is that low-rank matrices admit a much more succinct representation. It turns out that the same result holds for the Frobenius norm; i.e.,

$$A_{(k)} = \min_{B \in \mathbb{R}^{m \times n}, \operatorname{rank}(B) \leq k} \|A - B\|_F.$$  

(Recall that the Frobenius norm of a matrix is defined as \(\|C\|_F = \sqrt{\sum_{i,j} C_{i,j}^2}\).) Let’s see how this applies to our image compression problem.

Download the file `nash.jpg` into your Matlab path. You can read this file in by typing:

```matlab
1 A = imread('nash.jpg');
2 A = im2double(A);
3 A = rgb2gray(A);
```

The result is a 1500 × 981 matrix \(A\), with each entry representing a single pixel in the picture with a number between 0 and 1. To upload this picture on Instagram, you would need to upload 1500 × 981 = 1471500 numbers (pixels).

3. For \(k = 25, 50, 100, 200\), use Matlab to compute \(A_{(k)}\) as defined above. Report the value of \(\|A - A_{(k)}\|_F\) in each case. (Include your code for this part and the next.)

4. Use the commands `subplot` and `imshow` to produce on the same figure the original image, as well as your compressed images \(A_{(k)}\) for \(k = 25, 50, 100, 200\). Label your subplots. In addition, produce two separate plots demonstrating (i) \(\|A - A_{(k)}\|_F\) versus \(k\), and (ii) “total savings” versus \(k\). Total savings is to be interpreted as the answer to the question: How many fewer numbers do you need in order to store \(A_{(k)}\) than you did to store \(A\)? Explain why this number is equal to \(mn - (n + m + 1)k\). How much are you saving for \(k = 200\)?

5. Use the Matlab function `imwrite` to create two images from `imshow(A)` and `imshow(A_{(200)})`. Can you tell them apart? Does Nash get any less graceful?

**Problem 2: Local optimality and the gradient vector**

Let

$$f(x_1, x_2) = \frac{1}{2} x_1^2 + x_1 x_2 - \frac{3}{2} x_2^2 + 2x_1 + 5x_2 + \frac{1}{3} x_2^3.$$  

1. Find the local minimizers and maximizers of this function.
2. Without using arguments based on convexity, prove that for any quadratic function
\[ f(x) = x^TQx + b^Tx + c, \]
the following statements hold:

(a) \( \bar{x} \) is a local min \( \iff \nabla f(\bar{x}) = 0 \) and \( \nabla^2 f(\bar{x}) \succeq 0 \).

(b) \( \bar{x} \) is a strict local min \( \iff \nabla f(\bar{x}) = 0 \) and \( \nabla^2 f(\bar{x}) \succ 0 \).

Give counterexamples to show that these statements are not true in general.

3. The rate of increase of a function \( f : \mathbb{R}^n \to \mathbb{R} \) at point \( x \) and in a nonzero direction \( d \) is given by \( g'(0) \), where \( g : \mathbb{R} \to \mathbb{R} \) is defined as

\[ g(\alpha) = f \left( x + \alpha \frac{d}{\|d\|} \right). \]

Show that the minimum and maximum rates of increase are achieved in the directions \( -\nabla f(x) \) and \( \nabla f(x) \) respectively.

Problem 3: Norms, dual norms, and induced norms

1. Let \( Q \in \mathbb{S}^{n \times n} \) and assume \( Q \succ 0 \). Show that

\[ f(x) = \sqrt{x^TQx} \]

is a norm.

2. Show that \( Q^{-1} \) exists and is positive definite. Show that the dual norm of \( f \) is given by

\[ g(x) = \sqrt{x^TQ^{-1}x}. \]

(Hint: You may want to bring in \( \sqrt{Q} \), i.e., a matrix whose square is \( Q \). If you do, you have to first prove that this matrix exists.)

3. Let \( A \in \mathbb{R}^{m \times n} \). Prove the following expression for its induced 2-norm:

\[ \|A\|_2 = \sqrt{\lambda_{\text{max}}(A^TA)}. \]