(1) The SVD of a general, real-valued, rectangular matrix A can be recovered from the eigenvectors and eigenvalues of the augmented matrix:

$$K = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$$

Note that K is symmetric, so it does have a complete orthonormal basis of eigenvectors. Work out how to get the SVD of A from this basis. (Hint: take a look at K^2.)

(2) One possible way of getting the largest k singular values and corresponding singular vectors of A would be to use Orthogonal Iteration on K, with some random set of $2k$ initial vectors. Does this work? If not, suggest how to fix it. (Hint: as always, first try reducing a problem to the simplest and smallest case that reproduces the issue.) In addition to explaining your solution, code it up in MATLAB to illustrate that it works.

(3) In class (and on the lecture slides) the "optimal" step size for Steepest Descent was discussed, the α that minimizes $f(x+\alpha d)$ where $d = -\nabla f$ is the direction. This is, in some sense, what a line search algorithm is attempting to find; since Steepest Descent with this choice has poor convergence for ill-conditioned problems, we looked at choosing different directions (e.g. Newton's Method). However, an alternative is to take smarter step sizes, that may allow larger values of f in the short-term in the hope of improved convergence in the long-term. The Barzilai-Borwein (BB) method is just such an approach.

One derivation of BB is to think of it as a Quasi-Newton method where the Hessian is approximated as $H \approx \beta I$, a multiple of the identity matrix. Derive what β should be (based on the previous two iterations) according to the Quasi-Newton approximation property from class. Note there was a mistake in the lecture, the correct statement is:

$$H\Delta x \approx g(x + \Delta x) - g(x)$$

This is known as the "secant condition".

Implement this BB approach along with Steepest Descent, and test it on the quadratic diagonalized model problem:

$$f(x) = \frac{1}{2}\sum_{i=1}^{n} \lambda_i x_i^2$$

with $\lambda_1 \geq \cdots \geq \lambda_n > 0$. Recall the 2-norm condition number $\kappa(A) = \lambda_1/\lambda_n$. Compare the convergence of the two methods, particularly for large $\kappa(A)$.