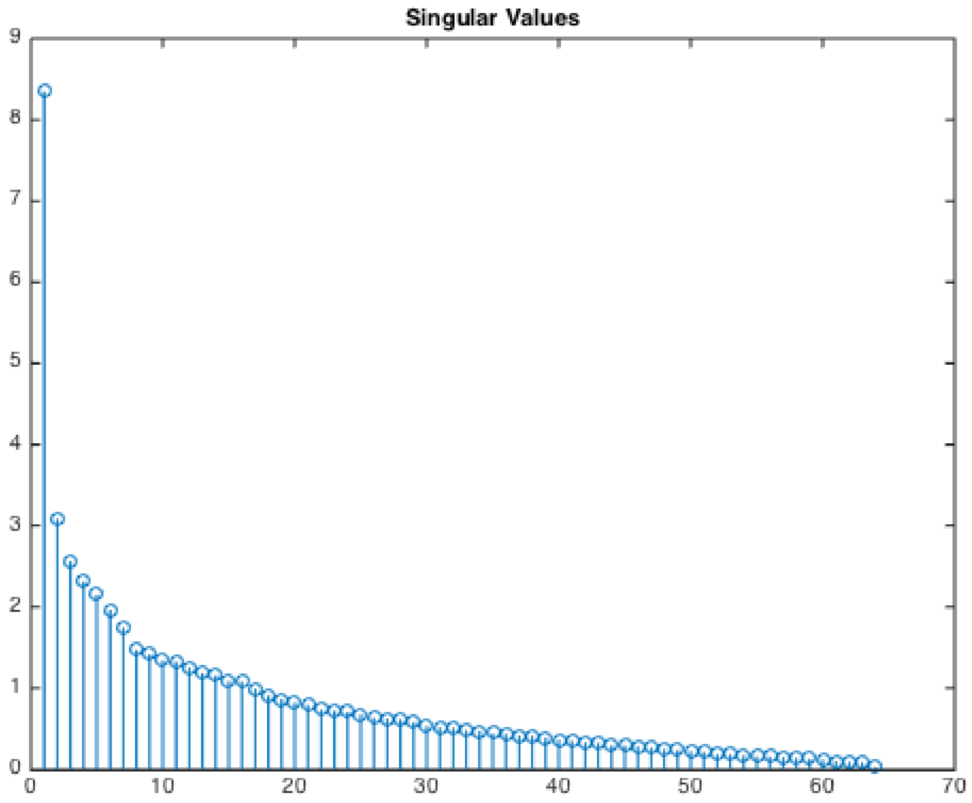
Contents

```
%15.39, 15.42, 16.16, 16.20, 16.23, 18.5
clear; close all;
term_by_doc;
```

Part A

```
for i = 1:n
   Ahat(:,i) = A(:,i)/norm(A(:,i));
end
[U,S,V] = svd(Ahat);
svalues = diag(S,0);
stem(svalues);
title('Singular Values');
```



Part B

```
q = zeros(64,1);
```

```
q(53) = 1;
results = Ahat'*q;
top = sort(results,'descend');
for i = 1:5
    top5(i) = find(results == top(i));
end

disp(['The top  5 results with query for the word students are: ' ...
        num2str(top5(1)) ', ' num2str(top5(2)) ', ' num2str(top5(3)) ', ' ...
        num2str(top5(4)) ', ' num2str(top5(5)) '.']);
```

```
The top  5 results with query for the word students are: 106, 105, 107, 115, 111.
```

Part C

```
% rank = 32
A8 = U(:,1:32)*S(1:32,1:32)*V(:,1:32)';
results = A8'*q;
top = sort(results,'descend');
for i = 1:5
    top5_32(i) = find(results == top(i));
end
disp(['RANK 32: The top  5 results with query for the word students are: ' ...
        num2str(top5_32(1)) ', ' num2str(top5_32(2)) ', ' num2str(top5_32(3)) ', ' ...
        num2str(top5_32(4)) ', ' num2str(top5_32(5)) '.']);


% rank = 16
A16 = U(:,1:16)*S(1:16,1:16)*V(:,1:16)';
results = A16'*q;
top = sort(results,'descend');
for i = 1:5
    top5_16(i) = find(results == top(i));
end
disp(['RANK 16: The top  5 results with query for the word students are: ' ...
        num2str(top5_16(1)) ', ' num2str(top5_16(2)) ', ' num2str(top5_16(3)) ', ' ...
        num2str(top5_16(4)) ', ' num2str(top5_16(5)) '.']);

% rank = 8
A8 = U(:,1:8)*S(1:8,1:8)*V(:,1:8)';
results = A8'*q;
top = sort(results,'descend');
for i = 1:5
    top5_8(i) = find(results == top(i));
end
disp(['RANK 8: The top  5 results with query for the word students are: ' ...
        num2str(top5_8(1)) ', ' num2str(top5_8(2)) ', ' num2str(top5_8(3)) ', ' ...
        num2str(top5_8(4)) ', ' num2str(top5_8(5)) '.']);

% rank = 4
A4 = U(:,1:4)*S(1:4,1:4)*V(:,1:4)';
results = A4'*q;
top = sort(results,'descend');
for i = 1:5
```

```
    top5_4(i) = find(results == top(i));
end
disp(['RANK 4: The top  5 results with query for the word students are: ' ...
        num2str(top5_4(1)) ', ' num2str(top5_4(2)) ', ' num2str(top5_4(3)) ', ' ...
        num2str(top5_4(4)) ', ' num2str(top5_4(5)) '.']);

%We can see that as we reduced the rank of the approximation the results of
%the query got worse and worse. By worse, I mean farther from the results
%that the full rank matrix provided. We can see however that with rank 32
%we got the exact same top 5 results as with the full rank case. With the
%rank 16 approximation the top result was still the same as the full rank
%query and was identical other than switching the 2nd and 3rd top query
%results. At ranks 8 and 4 the results got farther and farther from the
%best results.
```

```
RANK 32: The top  5 results with query for the word students are: 106, 105, 107, 115, 111.
RANK 16: The top  5 results with query for the word students are: 106, 107, 105, 115, 111.
RANK 8: The top  5 results with query for the word students are: 115, 106, 120, 107, 111.
RANK 4: The top  5 results with query for the word students are: 115, 105, 107, 66, 63.
```

Part D

```
%We can significantly reduce the cost of the search by using a lower rank
%approximation to the term-by-document matrix. In certain cases we saw
%similar or even identical results when using these low rank approxiamtions
%to the full rank search. Reducing the size of the term-by-document matrix
%results in significant savings in terms of computation time. Each
%reduction by a size of 2 reduces the computation time by a factor of 4 due
%to the quadratic nature of matrix multiplication.
```

Published with MATLAB® R2014b