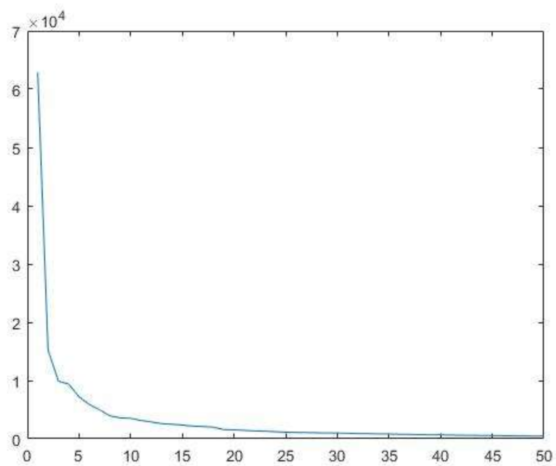
As massive amounts of data are produced each day, finding efficient ways of storing this data is critical. One method for approximating matrices which we will explore is through singular value decomposition. This method uses an approximation by the addition of the greatest n outer products where n is less than the overall rank of an image. Thus, the value of each singular value gives an indication of the weight of the outer product's effect on the final image. Since the outer product is given by each $\sigma_i u_i v^T$, the value of σ_i gives a scalar quantity of the weight of a particular outer product. In this investigation, the first 25, 50, and 100, as well as all 425 singular values were plotted on a graph of the i^{th} term verses its corresponding σ_i, thus linking the number of outer products used to the clarity of the image. As the following graph shows, after about the fifth term, the values for σ_i become about at least 10 times less than the first term. As we approach the 50^{th} term, the values for σ_i become almost zero when compared to the scale of the first singular value.

i^{th} term verses its corresponding σ_i, image 1



In fact, when the sum of the first 25 values of σ_i are taken as a percentage over the sum of the total values of σ_i, we find that these first 25 values account for 80.1% of the total variance in the picture. While this is a significantly high proportion of the total variance, greater variance is required in order to sharpen the image, as the image approximation for 25 singular values would be an unacceptable approximation due to its blurriness.

When considering an approximation for the first 50 outer products, we must consider that this approximation leads us to an image that, while a significant improvement upon the 25 singular value approximation, would still be an unacceptable approximation because the compression of the image is apparent in the uneven coloring in the gray background. This image contains about 89.6% of the total variance.

Lastly, the 100 value approximation shows the clearest image out of the approximations, containing 96.1% of the variance of the image. Since the 100 value approximation and the original image are almost indiscernible, this shows that the 100 value approximation is sufficient for successful image compression, reducing the number of outer product matrices by about 75%.

Next, the number of values it would take for the approximation to contain 99% of the variance of the image would be about 181 singular values. This implies that more than half of the outer products account for 1% of the images' variance, meaning that this half of the products contribute almost nothing to the approximation.

Note the choppiness of the below images where N= 25 and N = 50 and the indistinguishability between the images where N=100 and N=426.

N= 25

N= 50

N= 100

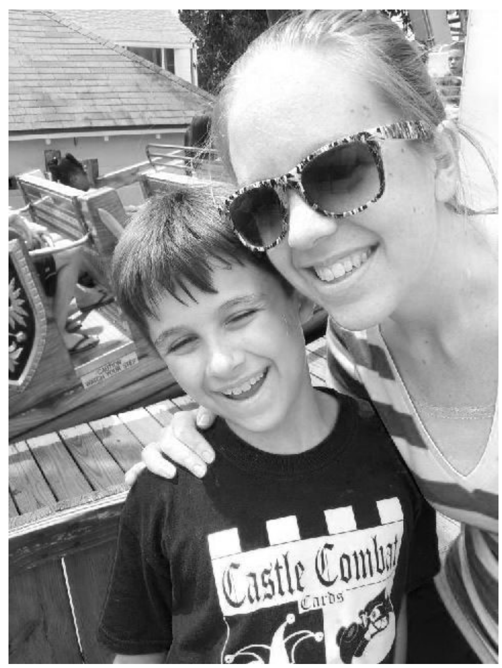
Original:
N= 426

Next, the same process is done with another picture.

N = 25

N = 50

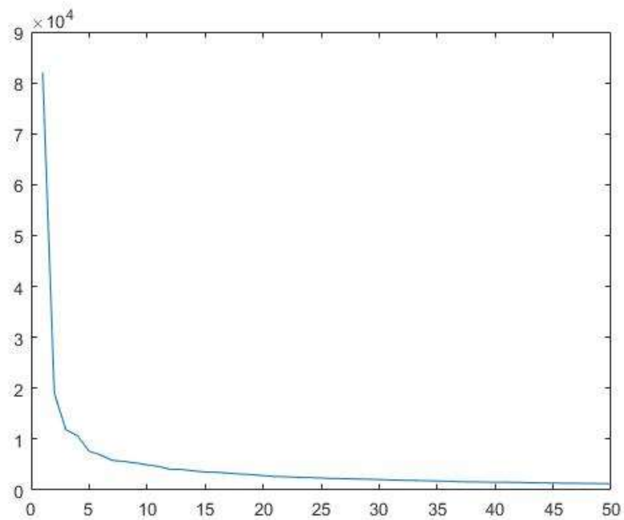
N = 100



Original:
N = 479

The graph of the SVDs of the second image is very similar to the graph of the SVDs of the first image. Both graphs drop off dramatically by about the fifth SVD, and show considerably smaller values by the 50th SVD value.

i^{th} term verses its corresponding σ_i, image 2



With the second image, however, the percentage of variance for 25, 50, and 100 SVD approximations is much lower. For this image, 25 SVDs account for 55.3% of the variance, 50 SVDs account for 66.5% variance, and 100 SVDs account for 78.2% of the variance. Additionally, in order for the approximation to account for 99% of the variance, it must sum about 397 outer products. Since there were 497 SVDs total for the second image, it was likely that more outer products were needed to account for the variance. Additionally, the second image may account for less of the variance per SVD because the second image had less area where the color was constant, and because the second picture had a larger original rank. The fact that 25, 50, and 100 SVDs for image 2 account for lower percentages of the variance than the SVDs for image 1 are evident in the fact that, overall, the approximations for the second image appear more grainy than the approximations for the first image. In particular, consider the approximation for N = 50 for the second image compared to the first image, and notice how the second image is much grainier than the first image for this value.

If each singular value is compared for N = 50, the first image's 50^{th} singular value = 449.1, while the second image's 50^{th} singular value = 1239.5, meaning that the weight of each singular value is more spread out over the distribution for the second image. Additionally, notice how N = 100 still isn't a satisfactory approximation for A in the second image, due to its graininess, while N = 100 is sufficient in approximating A for the first image.

In summary, since the second image requires significantly more SVDs to make an accurate approximation, the second image cannot be approximated as well as the first image using the singular value decomposition technique. This may be due to the fact that the second image is more complicated and somewhat larger than the first image. Thus, when rounded up, an approximation of about N = (¾)*rank(A) SVD values should suffice for images that have a rank between 426 and 479 and that have about this level of resolution.