AMATH 515: HWT 7

SUPER-INTELLIGENT AGENT

Applied Department, University of Washington, Seattle, WA
nghiiuu@uw.edu

1. Introduction

In this report, the Discrete Cosine Transform and the ℓ_1 Regression Problem are used to change the pixel space and recover the image "Son of Man" by Rene Magritte. A mask matrix will be used to mask missing data of the corrupted image. The Proximal Gradient Descent and CVS Library will be implemented to solve the regression problem.

2. Theory

(1) Derive an expression for $\text{prox}_{t\psi}(x)$ when $\psi(x) = \|x\|_1$. Show your work. The definition of $\text{prox}_{t\psi}(x)$ is the following:
Let $t > 0$,

$$\text{prox}_{t\psi}(x) = \arg\min_{x\in\mathbb{R}} \|z\|_1 + \frac{1}{2t}\|z - x\|_2^2$$

Consider the 2 following cases for the values of z_i for $i = 1...n$:
Denote $g(z) = \|z\|_1 + \frac{1}{2t}\|z - x\|_2^2$
Case 1: Suppose $z_i \neq 0$
Taking the gradient of $g(z)$ with respect to z component wise,

$$\nabla g(z) = \text{sgn}(z_i) + \frac{1}{t}(z_i - x_i)$$

To find the minimizer of $g(z)$, find z_i^* such that $\nabla g(z) = 0$,

$$0 = -\text{sgn}(z_i) + \frac{1}{t}(z_i - x_i)$$

(1)
$$z_i = \text{sgn}(z_i)t + x_i$$

Consider $z_i > 0$, then the sum 1 results to the following:

$$z_i = -t + x_i$$

is positive if $x_i > t$. Consider $z_i < 0$, then the sum 1:

$$z_i = -t + x_i$$

is negative if $x_i < -t$.

Case 2: Suppose $z_i = 0$
Note that the gradient of a L_1 norm, the first component of $g(z)$, is not defined as L_1 is not smooth and there does not exist one tangent line at $z = 0$.To handle the sharp corner,

Date: March 19, 2025.

examine the differential of $g(z)$ at $z=0$. The definition is as the following,
If $g \in \partial g(x)$ where

$$\begin{aligned}\partial g(x) =&\partial(\|z\|_1 + \frac{1}{2t}\|z-x\|_2^2)\\ &\partial(\|z\|_1) + \frac{1}{t}(z-x)\end{aligned}$$

, then g is the slope of a non-unique tangent line that touches $g(z)$ at some point z. Note that the sub-differential of $g(x)$ is additive as $g(z)$ is linear. For this case, we need to x_i such that the solution $z_i = 0$ of $\text{prox}_{t\psi}(x)$, in other words find x_i such that $0 \in \partial g(z)$. From lecture 13, $\partial(\|0\|_1) = \{g : g \in [-1,1]\}$. Then it follows that

$$\begin{aligned}0 \in&[-1,1] - \frac{x_i}{t}\\ \in&[-1-\frac{x}{t}, 1-\frac{x_i}{t}]\end{aligned}$$

Solving for x_i, we have $-1-\frac{x}{t} \leq 0$ implies $-\frac{x_i}{t} \leq 1$ $0 \leq 1-\frac{x}{t}$ implies $1 \leq 1-\frac{x_i}{t}$. It follows that, $x_i \in [-t,t]$.

Bose cases give rise to the following definition of

$\text{prox}_{t\psi}(x)$, $\begin{cases} x_i - t & \text{if } x > t \\ 0 & \text{if } x \in [-t,t] \\ x_i + t & \text{if } x < -t \end{cases}$

(2) Write $\nabla f(x)$ with A^{-1} and A where

$$f(x) = \frac{1}{2}\|M \odot (\text{iDCT}(X) - Y)\|_F^2$$

where X, Y and M are a matrices.

Using chain rule, denote the outer function as $g(X) = \|X\|_F^2$ and the inner function as $h(X) = M \odot (\text{iDCT}(X) - Y)$. Then the gradient of $g(X)$ is $\nabla g(X) = 2X$ and

$$h(x) = (\text{iDCT}^*(M \odot^* (M \odot ((\text{iDCT}(X) - Y))))$$

where * denotes the adjoint of each operator. To find the adjoint of $M\odot$ and iDCT, we will use the following definition: For an operator L and its adjoint L^*, $< L(x), y > = < x, L^*(y) >$ $\forall x, y$ in a finite dimension inner product space.
Expanding $< M \odot (X), Y >_F$ where F denotes the Frobenius norm,

$$\begin{aligned}< M \odot (X), Y >_F =& \sum_{i=0}^{N_y-1}\sum_{j=0}^{N_x-1} M_{ij}X_{ij}Y_{ij}\\ =& \sum_{i=0}^{N_y-1}\sum_{j=0}^{N_x-1} X_{ij}M_{ij}Y_{ij}\\ =&< X, M \odot (Y) >_F\end{aligned}$$

Since we are in a finite dimension inner product space, by definition, $M\odot^* = M\odot$. Note that since M is a masked matrix composed of ones and zeros, $M \odot M = M$.
Note, since we are given $< Y, \text{iDCT}(X) >_F = < \text{DCT}(Y), X >_F$, and the definition of $< Y, \text{iDCT}(X) >_F = < X, \text{iDCT}^*(Y) >_F$, this implies the adjoint of $\text{iDCT}^* = \text{DCT}$.
Thus,

$$\nabla f(x) = (\text{DCT}(M \odot ((\text{iDCT}(X) - Y)))$$

Now to get $\nabla f(x)$ in terms of A and A^-1, we need to vectorized $\nabla f(x)$ given that

$$\text{vec}(\text{DCT}(U)) = A\text{vec}(U), \quad \text{and} \quad \text{vec}(U) = A^{-1}\text{vec}((U)).$$

. Vectoring both sides,

$$\begin{aligned}\text{vec}(\nabla f(x)) &= \text{vec}((\text{DCT}(M \odot ((\text{iDCT}(X) - Y)))) \\ &= A(\mathbf{vec}(M) \odot (\text{vec}(\text{iDCT}(X)) - \text{vec}((Y)))) \quad = A(\text{vec}(M) \odot (A^{-1}\text{vec}(x) - \text{vec}(Y)))\end{aligned}$$

(3) Derive an upper bound for the Lipchitz constant of ∇f

Denote vec(X):=x, vec(Y):=y, vec(M), and vec(X) where $x, y, m, z \in \mathbb{R}^{Ny*Nx}$. First note the following properties of the Hadamard product:

- $a \odot b = b \odot a \ \forall \ a, b \ \in \mathbb{R}^n$

The gradient $\nabla f(x)$ is Lipchitz Continuous if $\exists \ L > 0$ such that $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - z\|$. Expanding the left hand side,

$$\begin{aligned}\|\nabla f(x) - \nabla f(z)\| &= \|A(m \odot (A^{-1}x - y)) - A(m \odot (A^{-1}z - y))\| \\ &= \|A(m \odot (A^{-1}x - y) - m \odot (A^{-1}z - y))\| \\ &= \|A(m \odot (A^{-1}x - y - A^{-1}z + y))\| \\ &= \|A(m \odot (A^{-1}x - A^{-1}z))\| \\ &= \|A(A^{-1}(x - y) \odot m)\| \\ &= \|m \odot (x - y)\|\end{aligned}$$

Note that since m is only composed of zeroes' and ones', when $m\odot$ is implied to $(x - y)$, some components of $x - y$ will be masked and some will remain the same. It follows that, $\|m \odot (x - y)\| < 1 * \|(x - y)\|$. This means we have found an upper bound of the some smallest Lipchitz constant.

3. Methods

The Discrete Cosine Transform (DCT) reconstructs a matrix representation of a corrupted image into a sum of cosine basis functions. DCT is similar to the Discrete Fourier Transform as both transformation represent images as frequencies, but DCT does not consider the complex part of a frequency. To recover missing data from an image, we need to solve the following ℓ_1 regression problem:

$$X^* = \text{argmin} \ \frac{1}{2}\|M \odot (\text{iDCT}(X) - Y)\|_2^2 + \lambda\|\text{vec}(X)\|_1$$

where X is the DCT of some matrix $\tilde{Y}$ that is similar to the true uncorrupted image, in other words, $\tilde{Y}$ is the iDCT of the solution X^*. Here, M is a masking matrix containing only zeros' and ones', when $M\odot$ is applied onto iDCT$(X) - Y$, we are minimizing the measure between the true data in the corrupted image Y and the approximating $\tilde{Y}$. In this paper, we are using implementing DCT and iDCT using Scipy and solving the regression problem with CVX and PGD (Proximal Gradient Descent).

4. Results

We used two different implementation to solve the ℓ_1 regression problem; Figure 1 is with CVX and Figure 2 is with the Proximal Gradient Descent for Down-sampled image to make the problem more simple. Both method results in a Mean Square Error of approximately .04. For figure 3, we used the Proximal Gradient Descent for a Full size image. For all 3 figures mentioned, lambda is.01 and the mask-proportion is .6. Figure 4 subfigure A shows the Mask and Corrupted image for the mask proportion being 1 and a fixed lambda .6. Both images are completely black as the mask proportion of 1 means the Mask matrix is completely zero and all of the data in corrupted image is masked. On the other hand, a mask proportion of .001, a number relatively small and close to zero, results in the Mask image to be completely white and the corrupted image remains mostly uncorrupted as a the matrix is mostly composed with ones. Figure 5 shows for a fixed mask proportion of .6, how does a larger or smaller lambda changes the reconstruction image. Figure A is where lambda is 1; the reconstruction is more blurry than the reconstruction when lambda is.6. The opposite happens as when lambda is .00001; as lambda gets smaller, then reconstruction gets closer to the true picture.

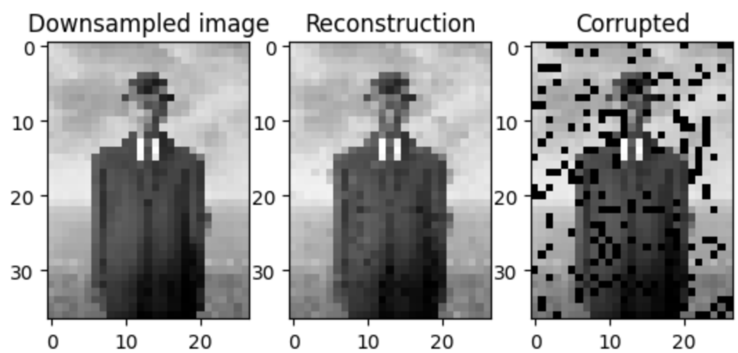


Figure 1. CVX Implementation for a Downsampled Image

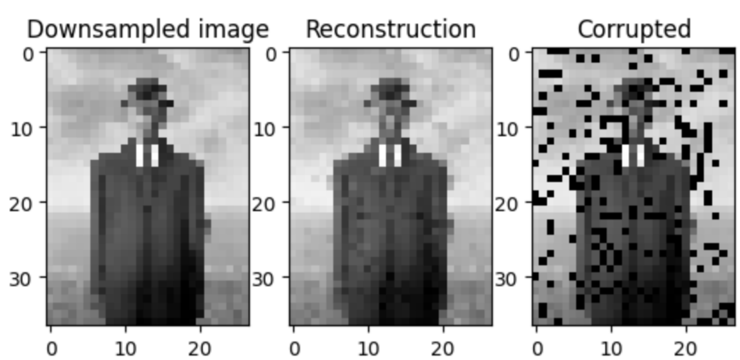


Figure 2. Proximal Gradient Descent Implementation for a Downsampled Image

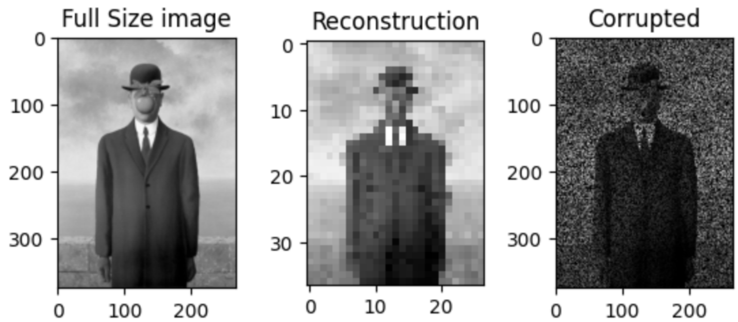
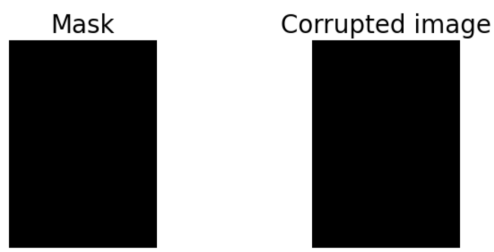


FIGURE 3. Proximal Gradient Descent Implementation for a Full Size Image

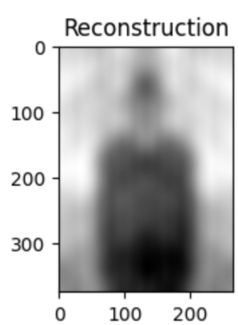
You may also need to include multiple figures:



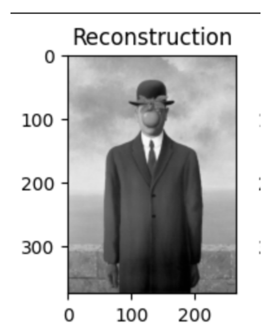
(A) First subfigure

(B) First subfigure

FIGURE 4. A shows the Mask and Corrupted Image for the mask proportion 1 and B is when the mask proportion is .0001 where lambda is fixed to .01



(A) First subfigure



(B) First subfigure

FIGURE 5. A shows the Reconstruction Image for $\lambda = 1$ 1 and B is when $\lambda = .00001$ where the mask proportion is fixed to .6

5. Summary and Conclusions

In conclusion, CVXPY and Proximal Gradient Descent find the solution of the most optimal reconstruction such that the regression problem is solved with the sparsity imposed by the DCT. Both give similar results for the down-sampled image. However, higher or lower combinations of lambda and mask proportion changes the reconstruction and corrupted picture. As lambda decreases, the reconstruction image gets closer to the true image, while a higher mask proportion masks more data of the corrupted image.

Acknowledgements

I am grateful for Alex Ballew, Cade Hsu, Johan Azambou, Crimson Stambaugh, and Emily Zhang.