Rensselaer 2020 REU Notebook

Alex Striff

May to July 2020

Contents

1 Project description

May 27, 2020

The aim of this REU project is to quantify the information present in images by the principled application of methods from statistical physics. The approach is to find a suitable notion of entropy which captures the salient features of particular kinds of images. We will consider a variety of features motivated by intuition or domain knowledge, and then move to machine learning as a tool for discovering other features.

2 Getting started

May 27, 2020

The initial goal is to characterize the most naïve calculation, which I'll call the *intensity entropy*. This does *not* take into account the spatial correlation of pixels in an image.

3 Intensity-level entropy

Given a discrete random variable X with support $\mathcal{X}$, the *Shannon entropy* is

$$H = \sum_{x \in \mathcal{X}} -P(x) \ln P(x).$$

The *intensity-level entropy* is the Shannon entropy of the empirical distribution of intensity values.

```
import numpy as np
```

```
def shannon_entropy(h):
    """The Shannon entropy in bits"""
    return -sum(p*np.log2(p) if p > 0 else 0 for p in h)

def intensity_distribution(data):
    """The intensity distribution of 8-bit `data`."""
    hist, _ = np.histogram(data, bins=range(256+1), density=True)
    return hist

def intensity_entropy(data):
    """The intensity-level entropy of 8-bit image data"""
    return shannon_entropy(intensity_distribution(data))

def intensity_expected(f, data):
    """The intensity-distribution expected value of `f`."""
    return sum(p*f(p) for p in intensity_distribution(data))
```

4 Effect of smoothing on intensity-level entropy

```
import numpy as np
import numpy.linalg as linalg
import matplotlib.pyplot as plt
from PIL import Image, ImageFilter, ImageOps
from src.utilities import *
from src.intensity_entropy import *
```

4.1 Natural image

```
img = ImageOps.grayscale(Image.open('test.jpg'))
scale = max(np.shape(img))
data = np.array(img)
img
```

```
intensity_entropy(img)
```

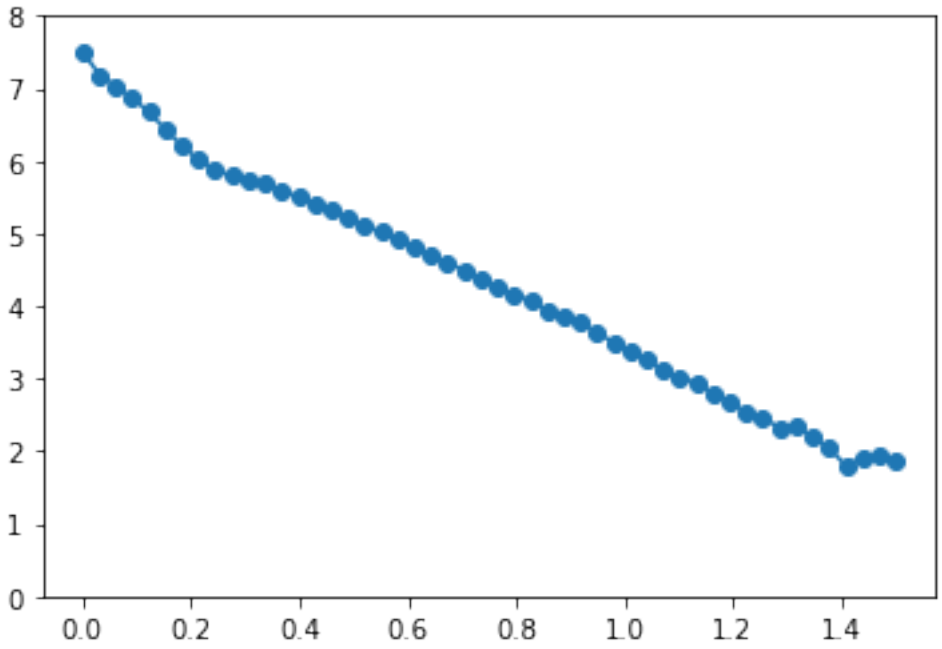
```
7.51132356216608
```

The problem with the intensity entropy is that it is usually near maximum (8 bits for these grayscale images).

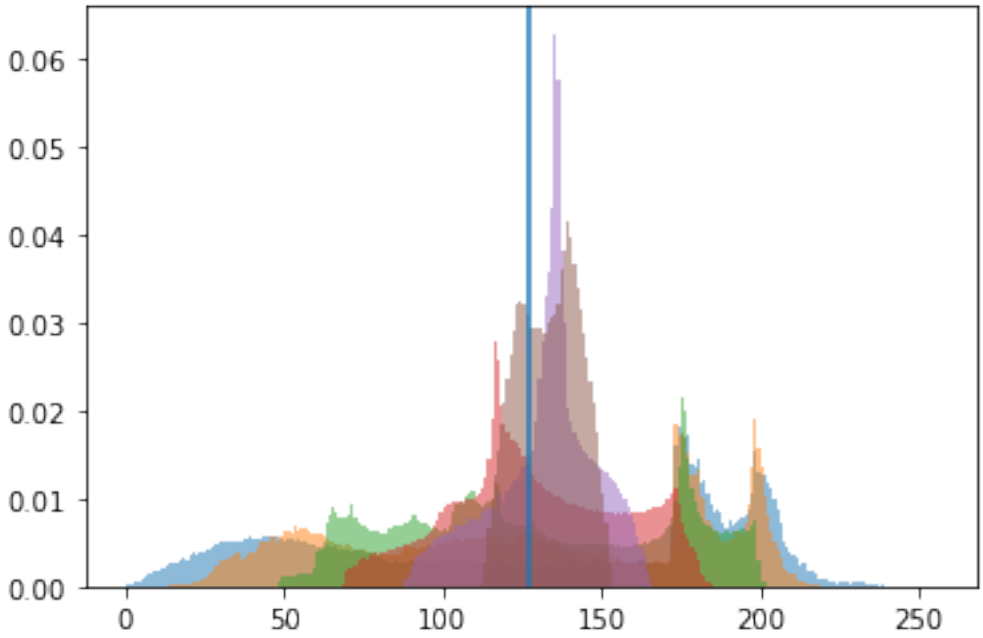
```
def intensity_blur(img, scales, display=True):
    scale = max(np.shape(img))

    results = []
    for k in scales:
        simg = img.filter(ImageFilter.GaussianBlur(k * scale))
        data = np.array(simg)
        ihist, ibins = np.histogram(data, bins=range(256+1), density=True)
        S = shannon_entropy(ihist)
        if display:
            hist = plt.hist(ibins[:-1], ibins, weights=ihist, alpha=0.5)
            results.append((k, simg, hist, S))
        else:
            results.append((k, S))

```

```
    if display:
        plt.axvline(x=np.mean(np.array(img)))

    return results
```

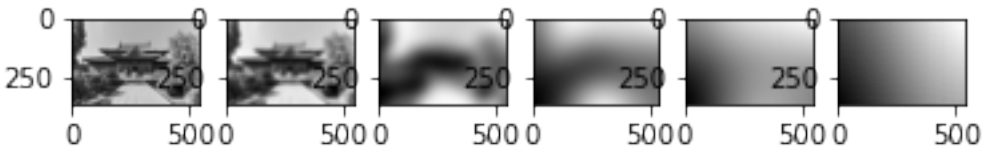
```
results = intensity_blur(img, np.linspace(0, 1.5, num=50), False)

plt.plot(*np.transpose(results), 'o-')
plt.ylim((0, 8))
plt.xlabel = "Smoothing"
plt.ylabel = "Intensity Entropy (bits)"
```




```
rimgs = [img for _, img, _, _ in intensity_blur(img, [0, 0.01, 0.05, 0.125, 0.25, 0.5])]
plt.show()
```

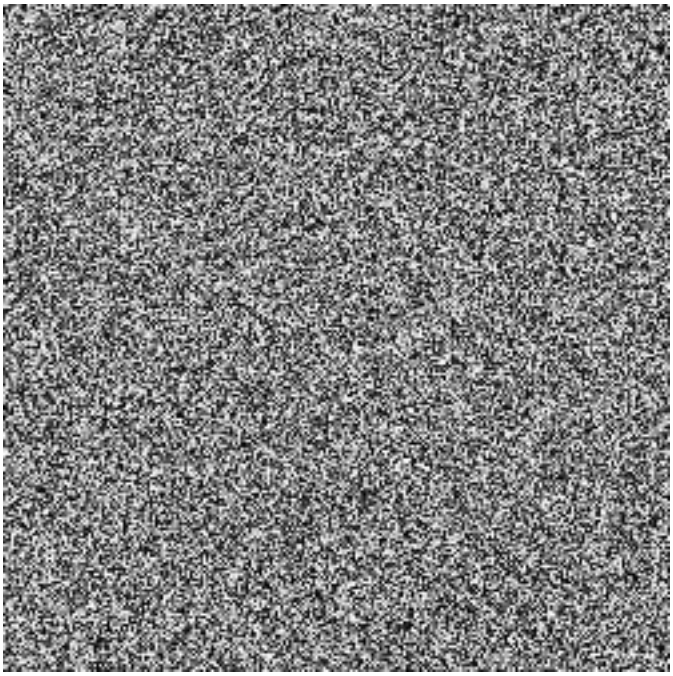



```
_, axarr = plt.subplots(1, len(rimgs))
for i, subimg in enumerate(rimgs):
    axarr[i].imshow(subimg, cmap='gray')
plt.show()
```



4.2 Random pixel values

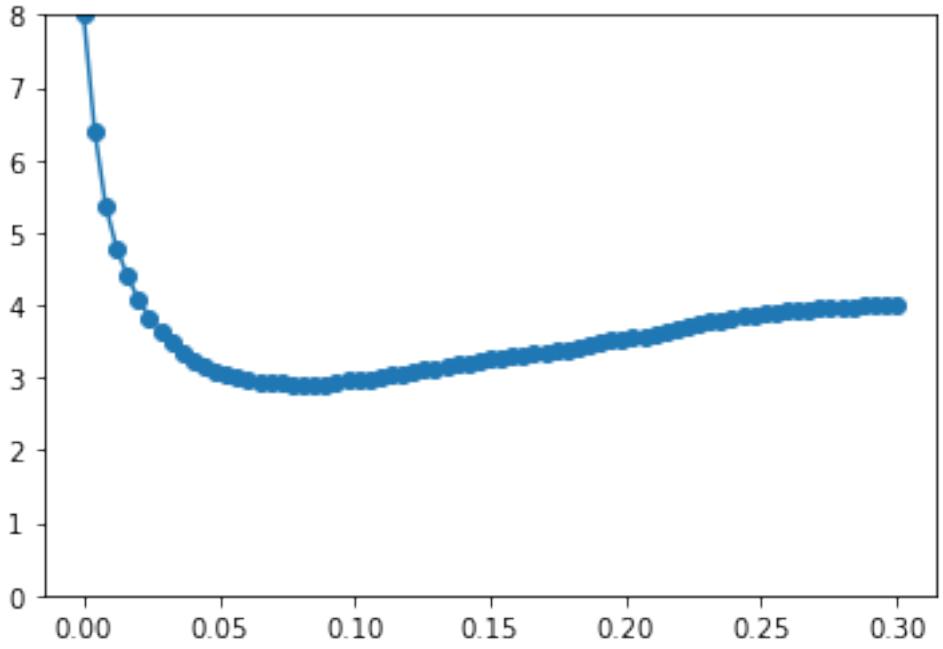
```
rsize = 250
randimg = Image.fromarray((256*np.random.rand(*2*[rsize])).astype('uint8'))
randimg
```



4.2.1 Beware: GIGO

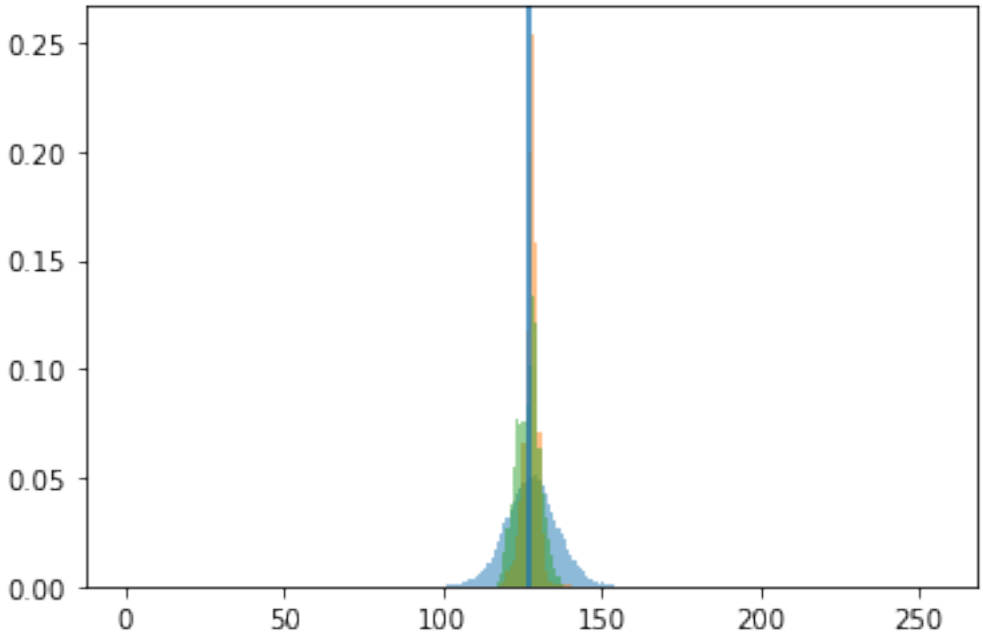
The boundary effects and discrete kernel of `ImageFilter.GaussianBlur` renders the data unreliable after the "minimum" of the intensity entropy with smoothing. This is immediately clear after even small smoothing for random pixel values, since there are no spatial correlations.

```
results = intensity_blur(randimg, np.linspace(0, 0.3, num=75), False)

plt.plot(*np.transpose(results), 'o-')
plt.ylim((0, 8))
plt.xlabel = "Smoothing"
plt.ylabel = "Intensity Entropy (bits)"
```

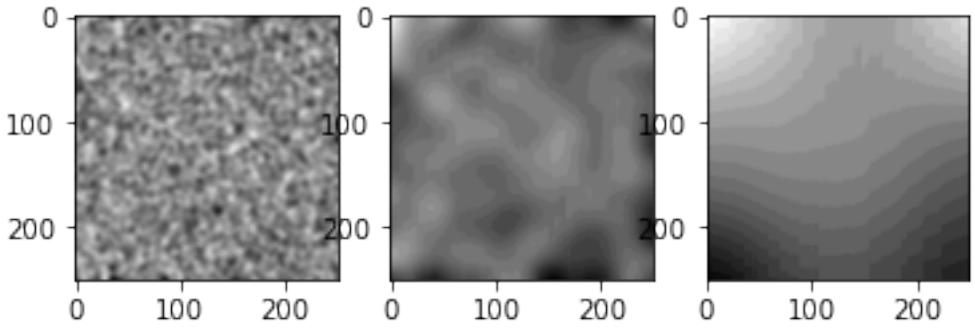



```
rimgs = [img for _, img, _, _ in intensity_blur(randimg, [0.01, 0.05, 0.25])]
```

```
plt.show()
```



```
_, axarr = plt.subplots(1, len(rimgs))
for i, subimg in enumerate(rimgs):
    axarr[i].imshow(subimg, cmap='gray')
plt.show()
```



The rightmost image should be uniform: the renormalization emphasizes incorrect deviations. These are what keep the intensity entropy from vanishing.

4.3 Comparing different levels of smoothing

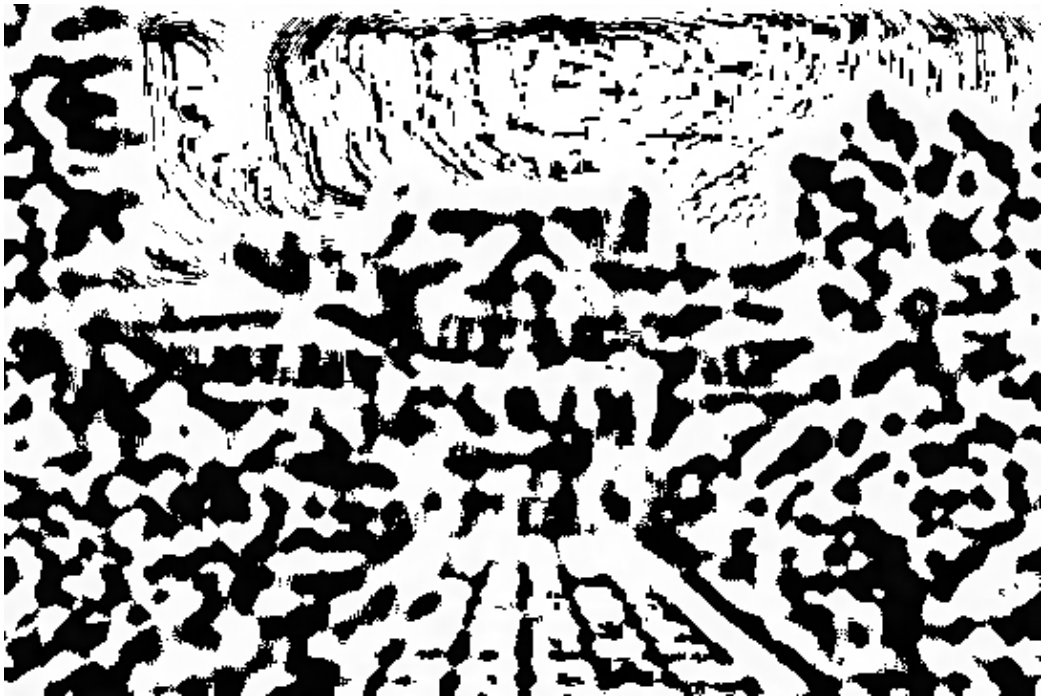
Is composing n Gaussian blurs with variance σ^2 the same as doing one with variance $n\sigma^2$ (considering the boundary effects and discrete kernel)?

```
nsmooths = 10
cimg = img
oneimg = cimg.filter(ImageFilter.GaussianBlur(np.sqrt(nsmooths)*2))
oneimg
```

```
nimg = cimg
for _ in range(nsmooths):
    nimg = nimg.filter(ImageFilter.GaussianBlur(2))
nimg
```

Answer: **No**
The differences between results at different scales can be pretty wack.

```
Image.fromarray((255*rescale(np.array(nimg) - np.array(oneimg))).astype('uint8'))
```



```
smimg = img
smdiff = np.array(smimg.filter(ImageFilter.GaussianBlur(2))) -
↪  np.array(smimg.filter(ImageFilter.GaussianBlur(100)))
diffimg = Image.fromarray((255 * rescale(smdiff)).astype('uint8'))
diffimg
```

5 Local metrics

May 28, 2020

Given an image $I : X \times Y \to \mathbb{Z}_n$, we will now consider *local metrics* for the information it contains.

I want to be careful in understanding the statistical assumptions I am making, so I'll try to be explicit about distinguishing true distributions from empirical distributions, and how the assumptions behind postulating the existence of empirical distributions relate to the actual calculation being done. This should also aid in learning more solid probability theory.

5.1 Induced metrics

Definition 1 (Lists). Given a set S, the collection of lists of elements from S is

$$\mathsf{List}(S) = \bigcup_{n \in \mathbb{Z}_{\geq 0}} S^n,$$

where a list (tuple) $s \in S^n$ is a map $s : \mathbb{Z}_n \to S$ and $|s| = n$.

Definition 2 (Image distributions). An *image distribution* is a map D that takes an image I and produces a random variable $D(I) : \Omega \to E$.

We are constructing empirical distributions from image data according to some map $M : \mathsf{Img} \to \mathsf{List}(\Omega)$, which produces the list of values $V = M(I)$. Then the probability of $D(I)$ taking a value in a subset $S \subseteq E$ is

$$P(X \in S) = \frac{1}{|V|} \sum_{s \in S} \left| V^{-1}(\{s\}) \right|.$$

Example 1. The intensity-level entropy is a function of the *nonnegative* random variable from the image distribution of intensity values. That is, the map M takes an image and returns the list of its intensity values.

Definition 3 (Induced image distributions). Given an image distribution D, and a subset $S \subseteq \operatorname{dom} I$, we construct the *induced image distribution* $D|_S$ by

$$D|_S(I) = D(I|_S).$$

Definition 4 (Induced random variable). Given an image I, an image distribution D and collection of subsets $\{S_i\}$ of $\operatorname{dom} I$, a function H admits the random variables

$$H_i = (H \circ D|_{S_i})(I)$$

Definition 5. The *r-box* at (x, y) is $B_r(x, y) = [x - r, x + r] \times [y - r, y + r]$.

Given two real random variables A and B with joint PDF $f_{A,B}(a, b)$, the PDF of their sum is

$$f_{A+B}(c) = \int_{-\infty}^{\infty} \mathrm{d}a \, f_{A,B}(a, a - c) = \int_{-\infty}^{\infty} \mathrm{d}b \, f_{A,B}(b - c, b). \tag{1}$$

For independent A and B, EQ. 1 reduces to $f_{A+B} = f_A * f_B$ over the marginals.

6 Kernels

Generalized to arbitrary functions on subregions of images.

```
1 import numpy as np
```

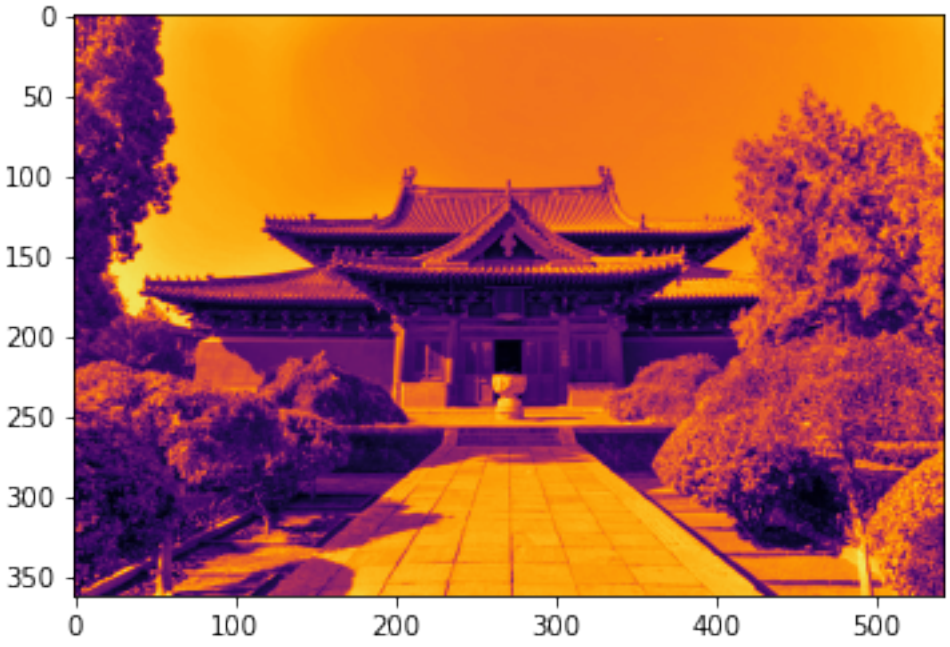
```
def box(x, y, r):
    return np.s_[max(0, x-r) : x+r+1, max(0, y-r) : y+r+1]
def mapbox(r, f, a):
    return np.reshape([f(a[box(*i, r)]) for i in np.ndindex(np.shape(a))], np.shape(a))
def mapboxes(rs, f, a):
    return (mapbox(r, f, a) for r in rs)
def mapallboxes(f, a):
    return mapboxes(range(max(np.shape(a))), f, a)

def mapblocks(h, w, f, a):
    return np.array([[f(y) for y in np.array_split(x, w, axis=1)]
                     for x in np.array_split(a, h)])
```

7 Boxcar intensity-level entropy

```
import numpy as np
import numpy.linalg as linalg
import matplotlib.pyplot as plt
from PIL import Image, ImageFilter, ImageOps
from src.utilities import *
from src.intensity_entropy import *
from src.kernels import *
plt.rcParams['image.cmap'] = 'inferno'
```

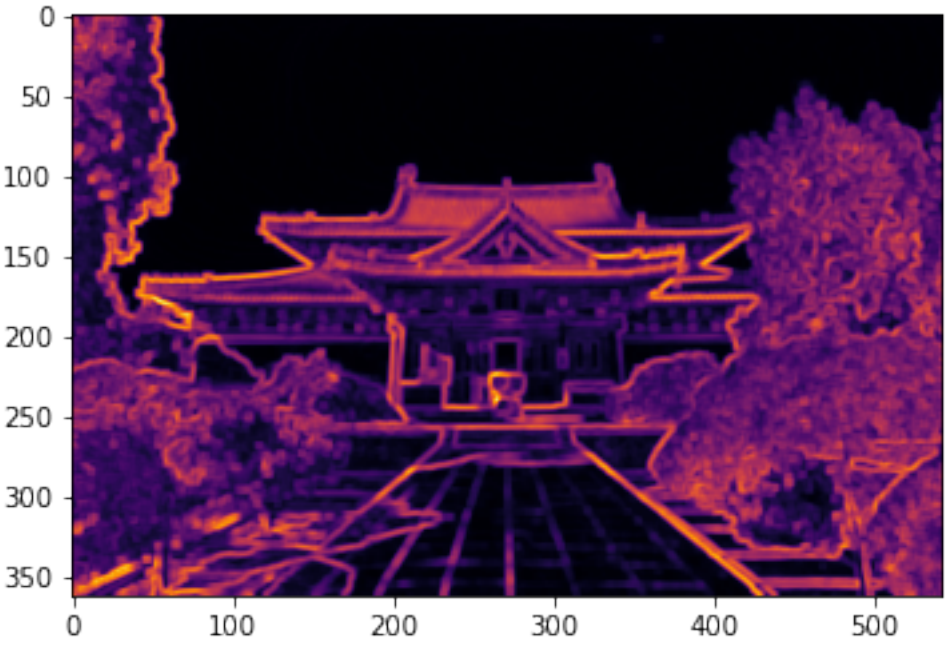
Let's compare the boxcar images for intensity entropy to those for a positive function on an image (the standard deviation) and for different functions of the induced intensity distribution.

```
img = ImageOps.grayscale(Image.open('test.jpg'))
scale = max(np.shape(img))
data = np.array(img)
plt.imshow(img);
```



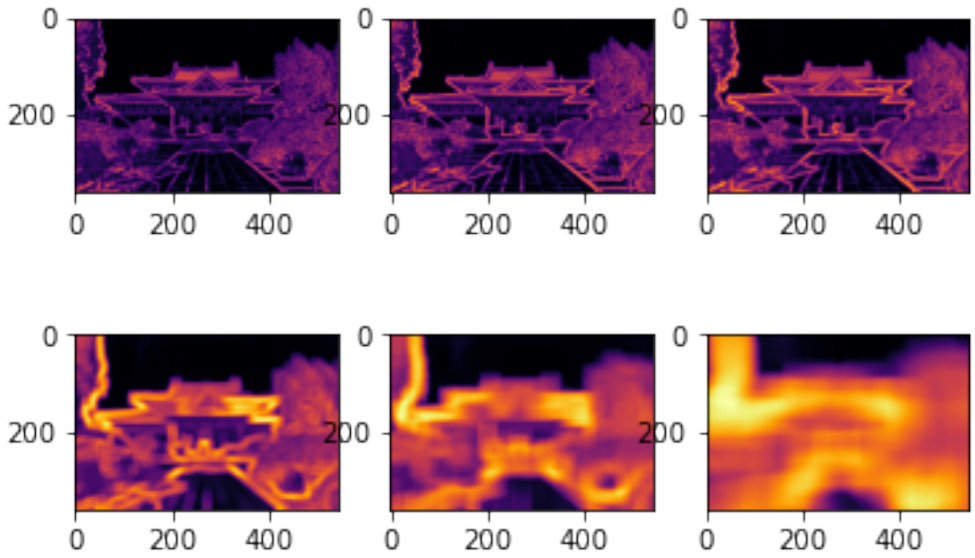
7.1 Standard deviation

```
plt.imshow(mapbox(2, np.std, np.array(img)));
```



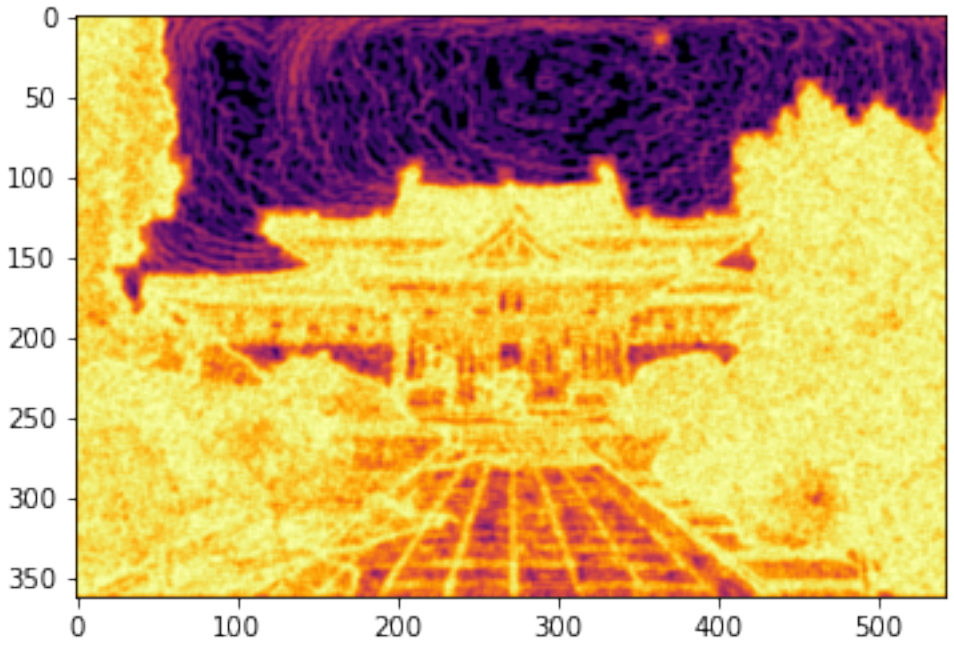

```
boxos = list(mapboxes([1,2,3,10,20,50], np.std, np.array(img)))
```

```
_, axarr = plt.subplots(2, np.ceil(len(boxos)/2).astype('int'))
for i, subimg in enumerate(boxos[:3]):
    axarr[0,i].imshow(subimg)
for i, subimg in enumerate(boxos[3:]):
    axarr[1,i].imshow(subimg)
plt.show()
```



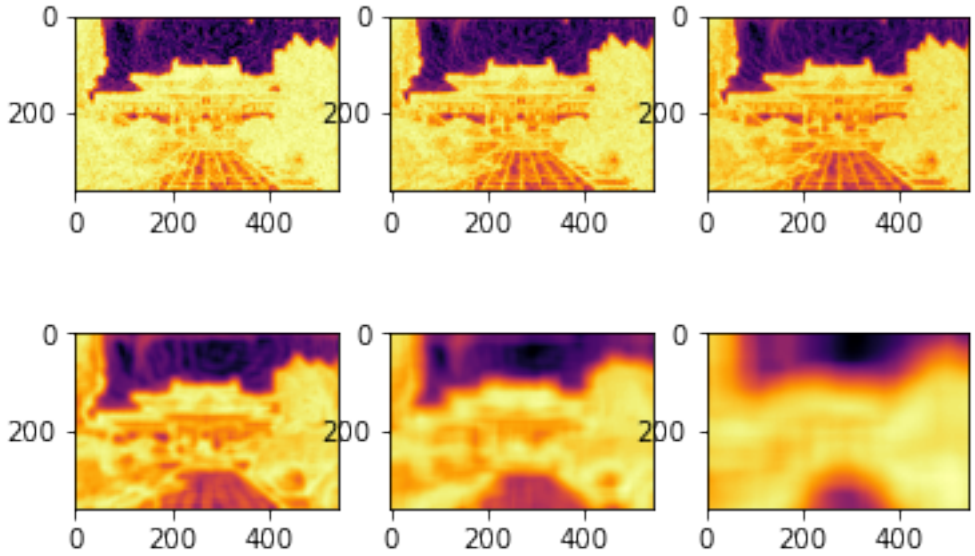
7.2 Intensity entropy

```
plt.imshow(mapbox(2, intensity_entropy, np.array(img)));
```



```
boxSes = list(mapboxes([1,2,3,10,20,50], intensity_entropy, np.array(img)))
```

```
_, axarr = plt.subplots(2, np.ceil(len(boxSes)/2).astype('int'))
for i, subimg in enumerate(boxSes[:3]):
    axarr[0,i].imshow(subimg)
for i, subimg in enumerate(boxSes[3:]):
    axarr[1,i].imshow(subimg)
plt.show()
```

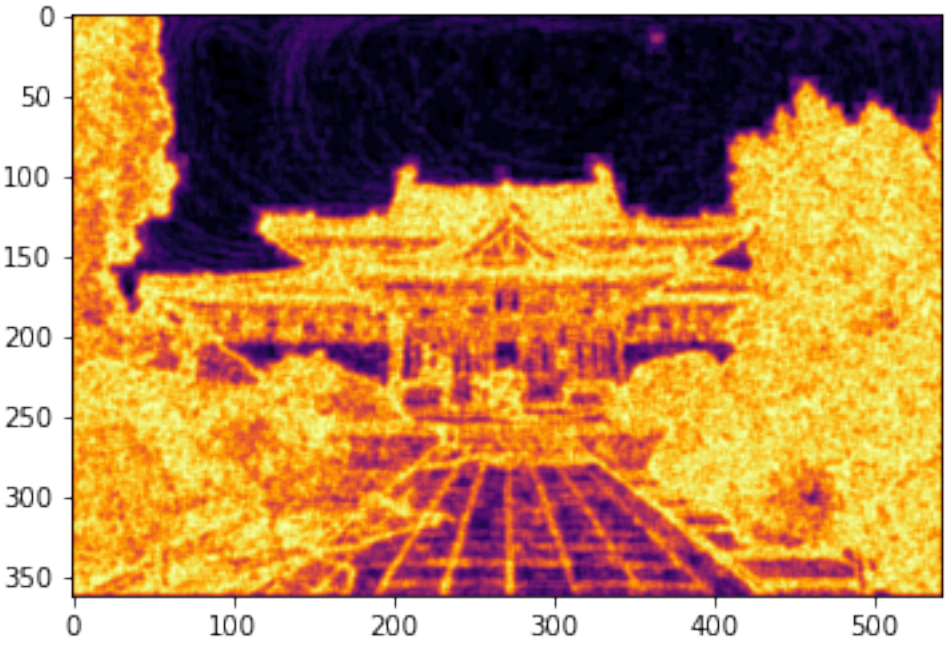


7.3 Replace surprisal with other functions

To what extent do the surprisal-related results depend upon the specific form of the *surprisal* $x \mapsto -\log(x)$ in the expected value of the intensity distribution? We will replace the expected surprisal with the expected f, for different functions f on the empirical probabilities of a pixel taking some intensity.

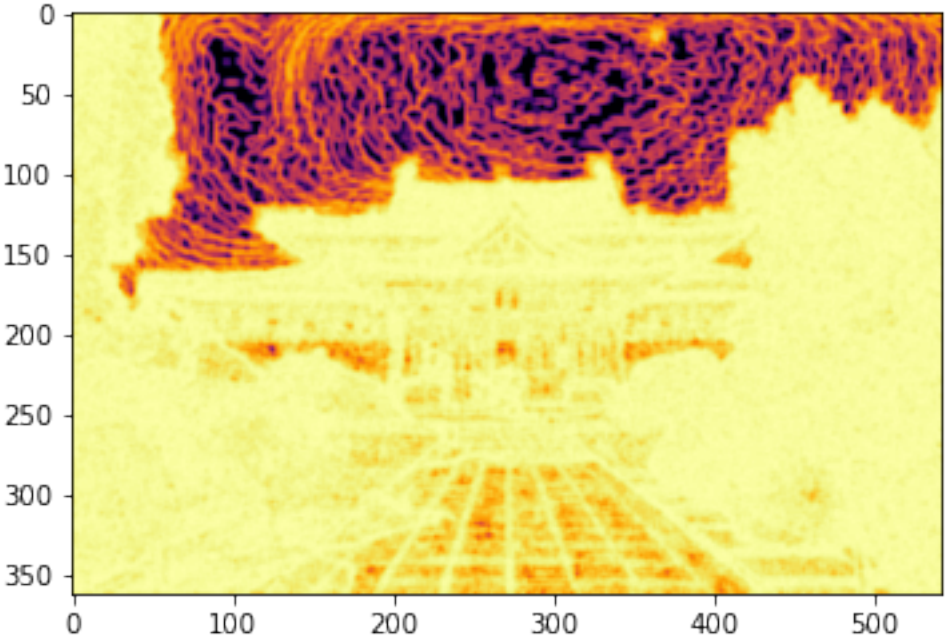
Laurent: $p \mapsto -1 + 1/p$.

```
plt.imshow(mapbox(2, lambda I: intensity_expected(lambda p: -1 + 1/p if p > 0 else 0, I),
↪  np.array(img)));
```



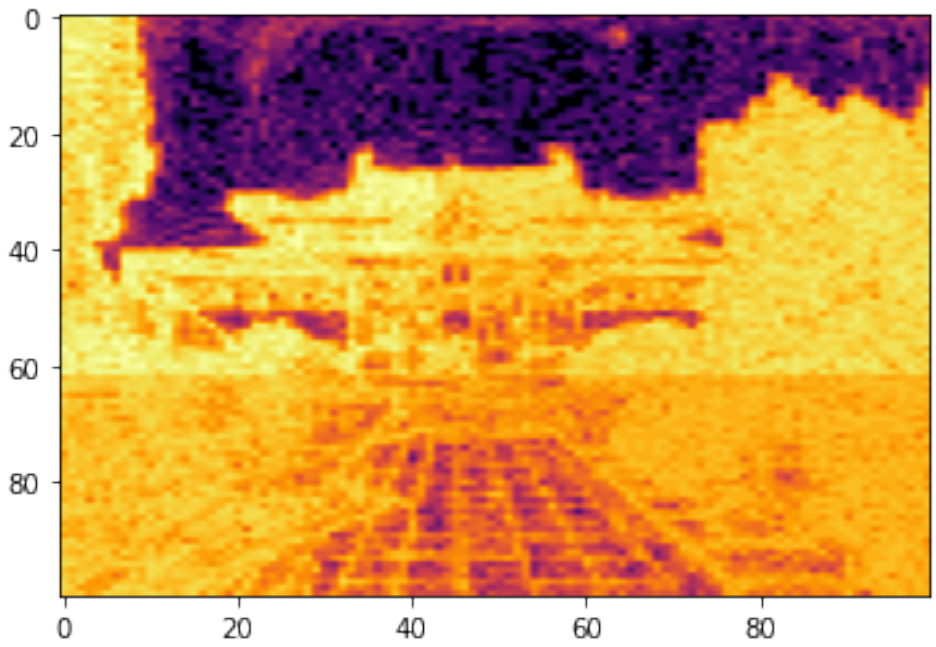
Taylor: $p \mapsto -(1+p)$.

```
plt.imshow(mapbox(2, lambda I: intensity_expected(lambda p: -(1+p), I), np.array(img)));
```

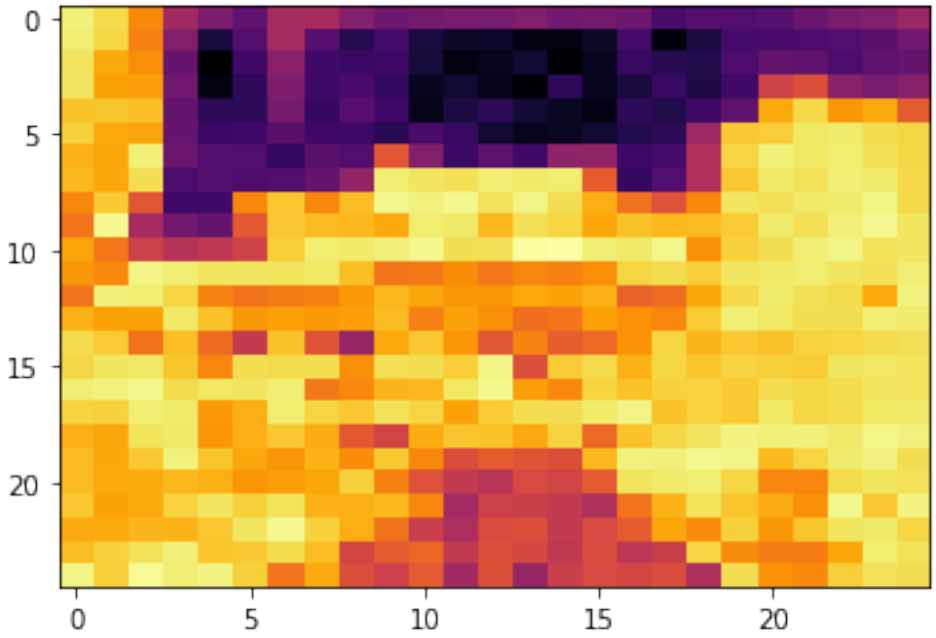


7.4 Intensity entropy on disjoint blocks

```
plt.imshow(mapblocks(100, 100, intensity_entropy, np.array(img)),
           aspect=np.divide(*np.shape(img)));
```




```
plt.imshow(mapblocks(25, 25, intensity_entropy, np.array(img)),
           aspect=np.divide(*np.shape(img)));
```



8 Fractal dimensions

May 29, 2020

The previous results hint at characterizing the growth of the intensity entropy with different discretizations.

Definition 6. The *Rényi entropy of order* $\alpha \geq 0$ of a discrete random variable X with support $\mathcal{X}$ is

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P(x)^\alpha = \frac{\alpha}{1-\alpha} \log \|P\|_\alpha,$$

where $\|P\|_\alpha$ denotes the α-norm of the vector of probability values. The limit $\alpha \to 1$ reproduces the Shannon entropy.

Definition 7. Given a real random variable X, define a discretized random variable

$$\langle X \rangle_\varepsilon = \frac{\lfloor \varepsilon X \rfloor}{\varepsilon}.$$

Then the *generalized dimension* of X is

$$d_\alpha(X) = \lim_{\varepsilon \to 0} \frac{H_\alpha(\langle X \rangle_\varepsilon)}{\log \varepsilon} = \lim_{\varepsilon \to 0} \frac{\alpha}{1-\alpha} \log \left(\|\langle X \rangle_\varepsilon\|_\alpha - \varepsilon \right).$$

The case $\alpha \to 1$ is the *information dimension* of X. The generalized dimension may be estimated from linear regression of $H_\alpha(\langle X \rangle_\varepsilon)$ with $\log \varepsilon$ as the independent variable.

9 Fractal dimension regression

```
import numpy as np
import numpy.linalg as linalg
import matplotlib.pyplot as plt
from PIL import Image, ImageFilter, ImageOps
from scipy import interpolate
from scipy import integrate
from src.intensity_entropy import *
from src.kernels import *
plt.rcParams['image.cmap'] = 'inferno'
```

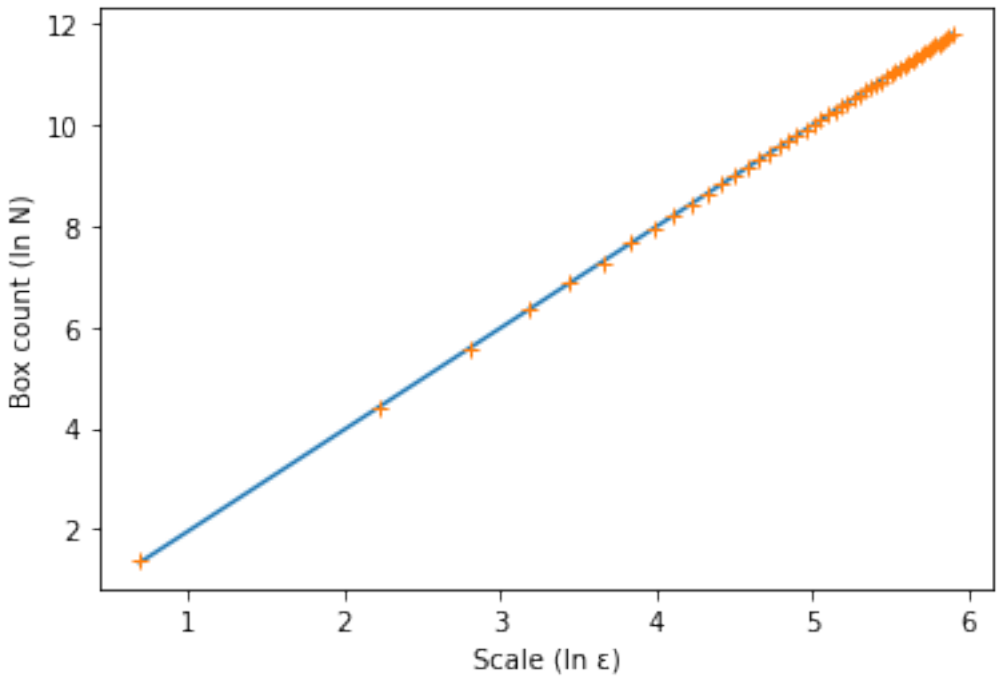
```
img = ImageOps.grayscale(Image.open('test.jpg'))
scale = max(np.shape(img))
data = np.array(img)
img
```

9.1 Box-counting dimension

```
def boxdim(data):
    εs = np.linspace(2, min(np.shape(data)))
    boxes = [np.log(np.sum(mapblocks(
        ε, ε, lambda x: 1 if np.any(x) else 0, data))) for ε in εs]
    logεs = np.log(εs)
    endεs = logεs[[0, -1]]
    dimfit = np.polyfit(np.log(εs), boxes, 1) # [slope, intercept]
    plt.plot(endεs, dimfit[0]*endεs + dimfit[1])
    plt.plot(logεs, boxes, '+')
    plt.xlabel('Scale (ln ε)')
    plt.ylabel('Box count (ln N)')
    return dimfit[0]
```

```
boxdim(data)
```

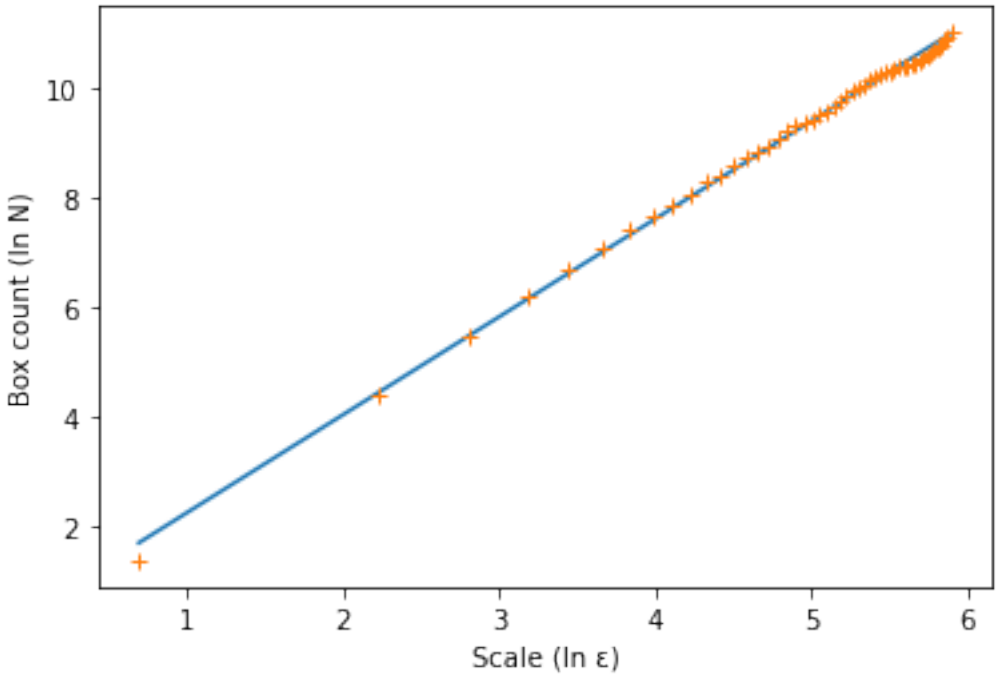
```
2.0087040269581435
```



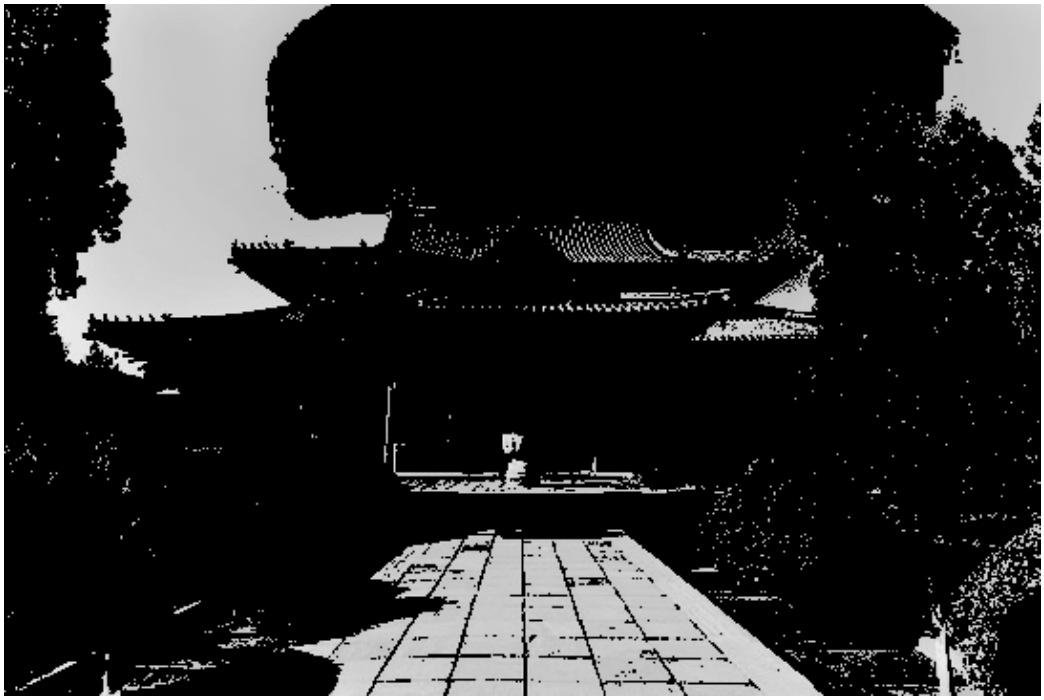

```
sky = data.copy()
sky[sky < 128+32] = 0
Image.fromarray(sky)
```

```
boxdim(sky)
```

```
1.7877778191348215
```

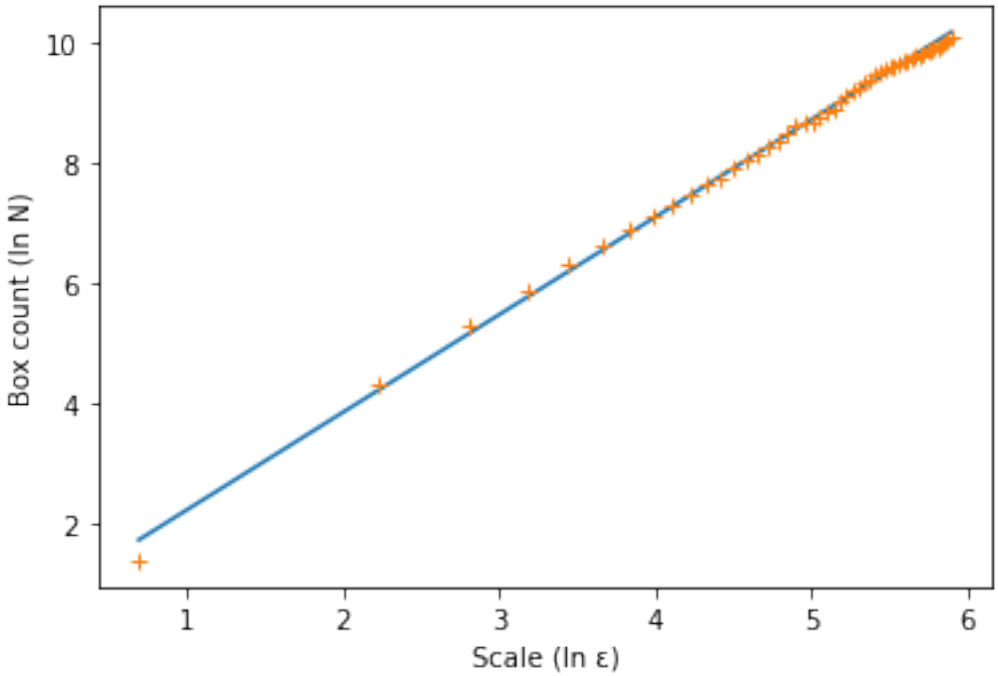



```
nosky = data.copy()
nosky[nosky < 128+64] = 0
Image.fromarray(nosky)
```



```
boxdim(nosky)
```

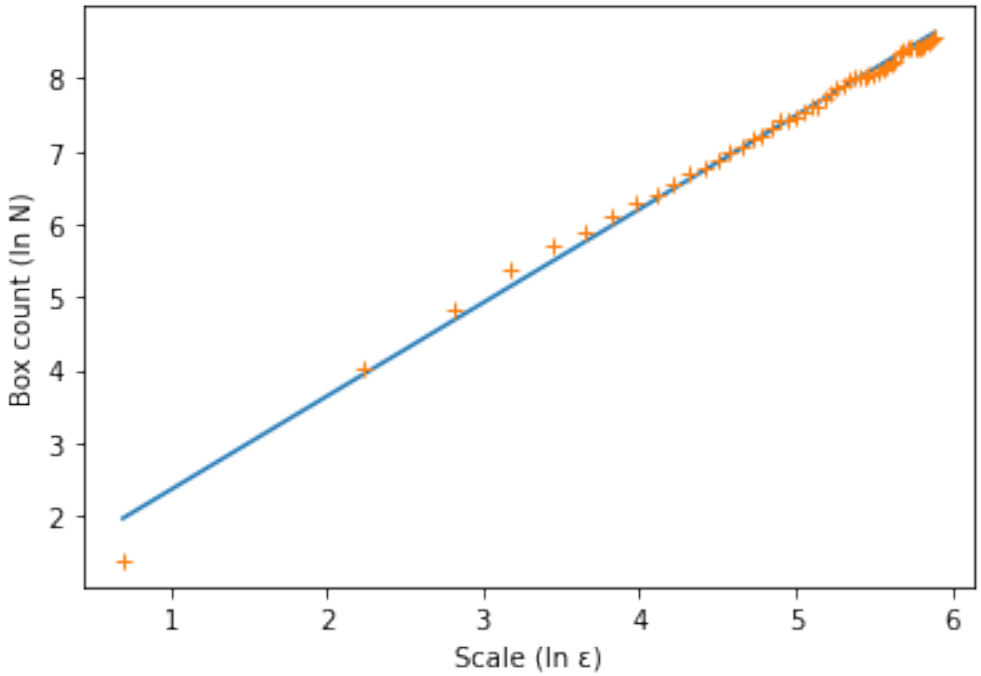
```
1.6214794967487127
```




```
dots = data.copy()
dots[nosky < 128+64+16] = 0
Image.fromarray(dots)
```

```
boxdim(dots)
```

```
1.2821025677557252
```



9.2 Information dimension

```
def discretize(f, a, b, ε, N=20):
    return [integrate.simps(f(np.linspace(c - ε/2, c + ε/2, N)), dx=ε / (N - 1))
            for c in np.arange(a + ε/2, b, ε)]
```

```
def infodim(dist, s=1e-5):
    l = len(dist)
    spl = interpolate.splrep(range(l), dist, s=s)
    f = lambda x: interpolate.splev(x, spl)

    εs = l / np.linspace(10, l)
    logεs = -np.log2(εs)
    endεs = logεs[[0, -1]]
    entropies = [shannon_entropy(discretize(f, 0, l, ε)) for ε in εs]
    dimfit, cov = np.polyfit(logεs, entropies, 1, cov='unscaled')

    plt.plot(endεs, dimfit[0]*endεs + dimfit[1])
    plt.plot(logεs, entropies, '+')
    plt.xlabel('Scale (lg ε)')
    plt.ylabel('Shannon entropy (bits)')
```

```
16
17      return dimfit[0], cov[0,0]
```

The Gaussian distribution

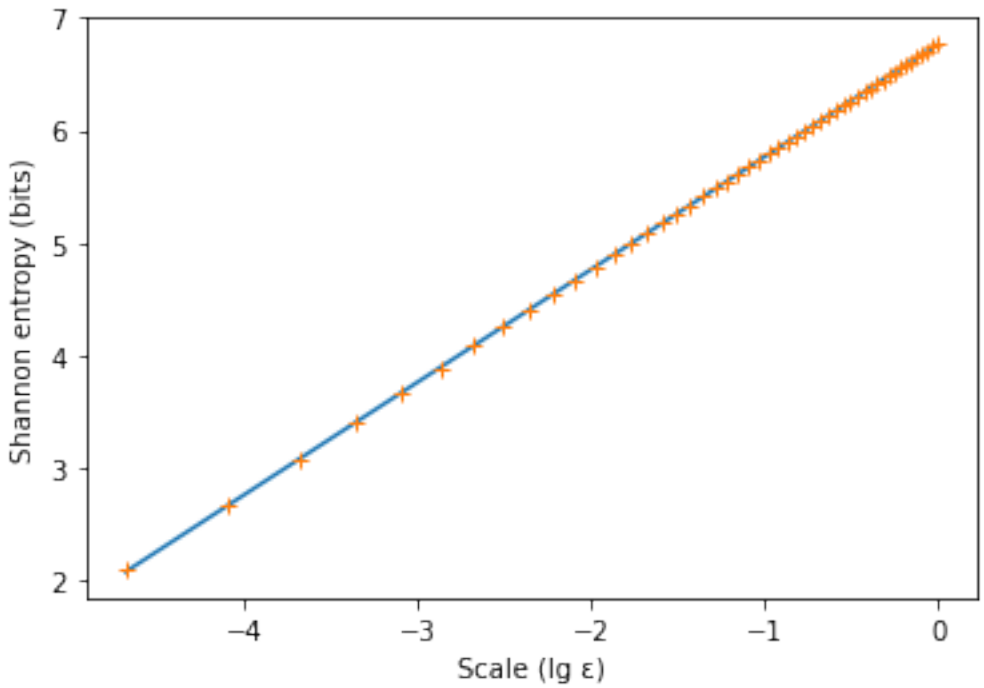
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

is continuous, so its information dimension is 1.

```
1  def gaussian(μ, σ, x):
2      return np.exp(-(x - μ)**2 / (2*σ**2)) / (σ*np.sqrt(2*np.pi))
```

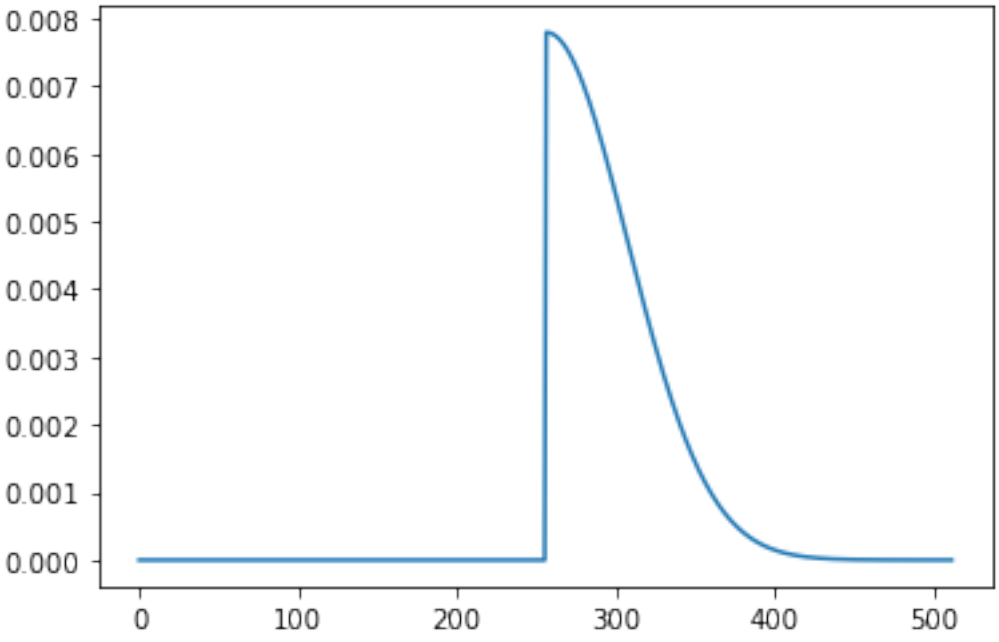
```
1  infodim((10/256) * gaussian(0, 1, np.linspace(-5, 5, 256)))
```

```
(1.0014184290221988, 0.015707497682893923)
```



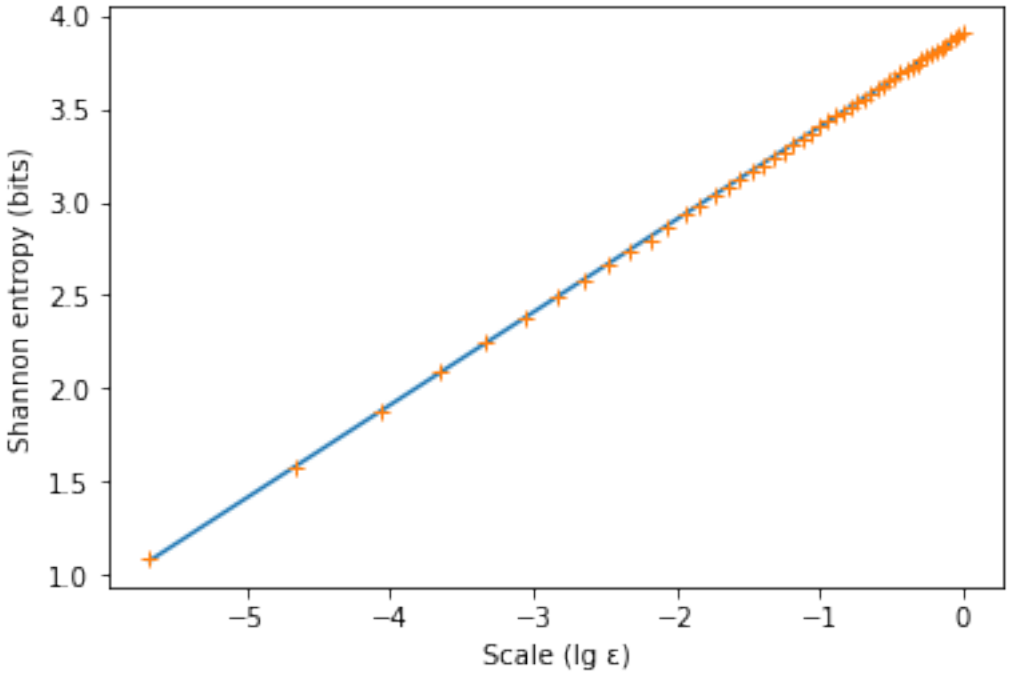
The rectified Gaussian distribution $g(x) = \Theta(x)f(x)+\delta(x)/2$ is half-continuous, so its information dimension is 1/2.

```
1  dist = np.concatenate([[0]*256, (5/256)*gaussian(0, 1, np.linspace(0, 5, 256))])
2  plt.plot(dist);
```



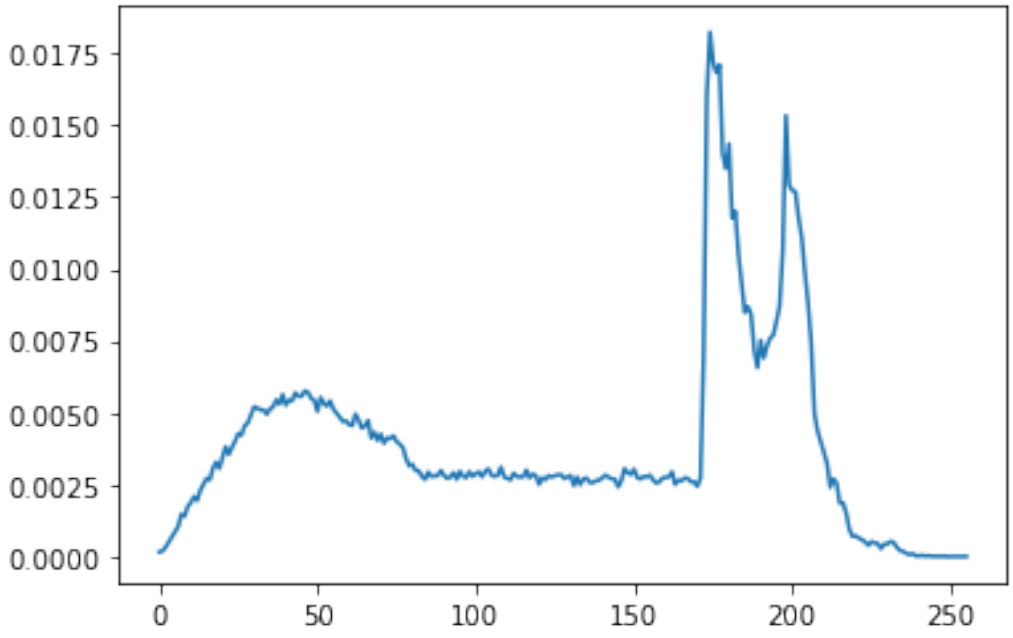

```
infodim(dist)
```

```
(0.4979088715795226, 0.012313889825394398)
```



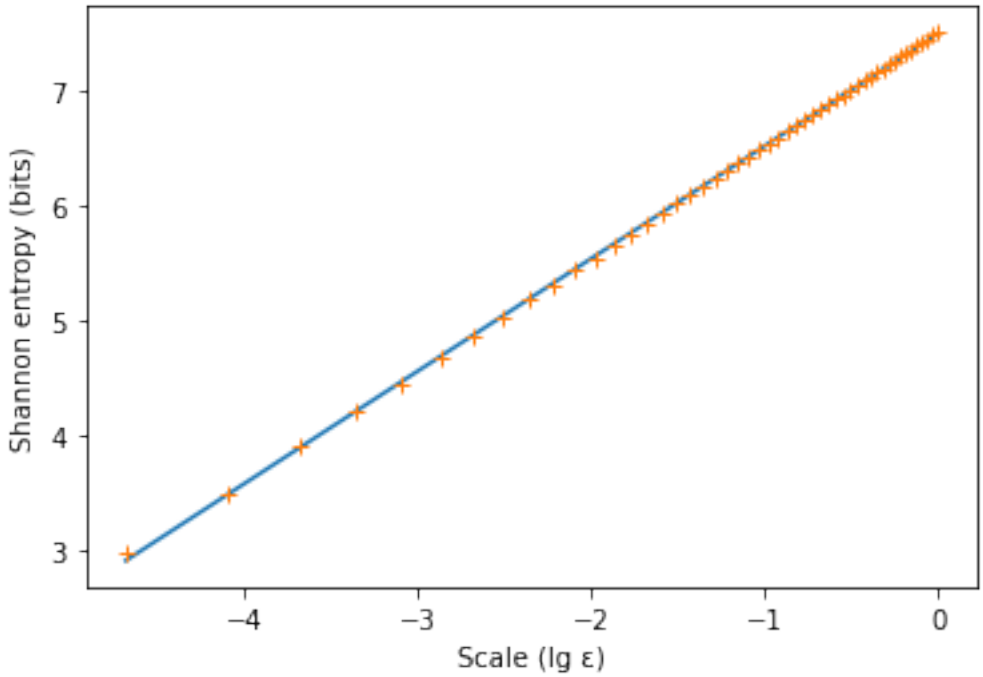
Now that we've validated `infodim`, what does it say about the intensity distribution of an image?

```
dist = intensity_distribution(img)
plt.plot(dist);
```




```
infodim(dist)
```

```
(0.98265843047326, 0.015707497682893923)
```



10 Probability and inference

May 30, 2020

Let's look at a simple inference problem before considering images. This example illustrates how the approach founded on probability theory differs from the naïve statistical approach usually taken by physicists.

Example 2 (Biased coin tosses). Consider tossing a biased coin N times to obtain n heads. What is the probability p' that the next coin toss comes up heads?

The temptation is to claim n/N as the probability, but this is *incorrect* if we want to allow all consistent biases. The problem with this solution is that the most probable bias is assumed to be the true bias.

The probability of getting m heads if a single head has probability p is

$$P(m \mid p) = \binom{N}{m} p^m (1-p)^{N-m}.$$

We have no other information, so we assume that all of the biases are equally likely. This means that $P(p)$ is constant (the uniform prior). The distribution of

biases p given the observation of m heads is then

$$P(p \mid m) = \frac{P(m \mid p)P(p)}{P(m)} = \frac{P(m \mid p)P(p)}{\int_0^1 \mathrm{d}\tilde{p}\, P(m \mid \tilde{p})P(\tilde{p})} = \frac{P(m \mid p)}{\int_0^1 \mathrm{d}\tilde{p}\, P(m \mid \tilde{p})}.$$

We compute that

$$P(m) = \binom{N}{m} \int_0^1 \mathrm{d}p\, p^m (1-p)^{N-m} = \binom{N}{m} \frac{m!(N-m)!}{(N+1)!} = \frac{1}{N+1},$$

so the next coin toss is heads with probability

$$\begin{aligned} p' &= \int_0^1 \mathrm{d}p\, P(\text{head} \mid n,\, p)P(p \mid n) = \int_0^1 \mathrm{d}p\, p\, P(p \mid n) \\ &= \int_0^1 \mathrm{d}p\, p(N+1)\binom{N}{n} p^n (1-p)^{N-n} = \frac{n+1}{N+2}. \end{aligned}$$

For $n = 3$ and $N = 10$, $p' = 0.33$. This is a more conservative estimate than $p' = 0.30$ from the most probable bias.

11 Ising images

June 1, 2020 What happens if we apply a model from statistical physics to an image?

12 Ising images

```
import numpy as np
import numpy.linalg as linalg
import matplotlib.pyplot as plt
from PIL import Image, ImageFilter, ImageOps
import imageio
plt.rcParams['image.cmap'] = 'gray'
```

```
from ipywidgets import IntProgress
from IPython.display import display
import time
```

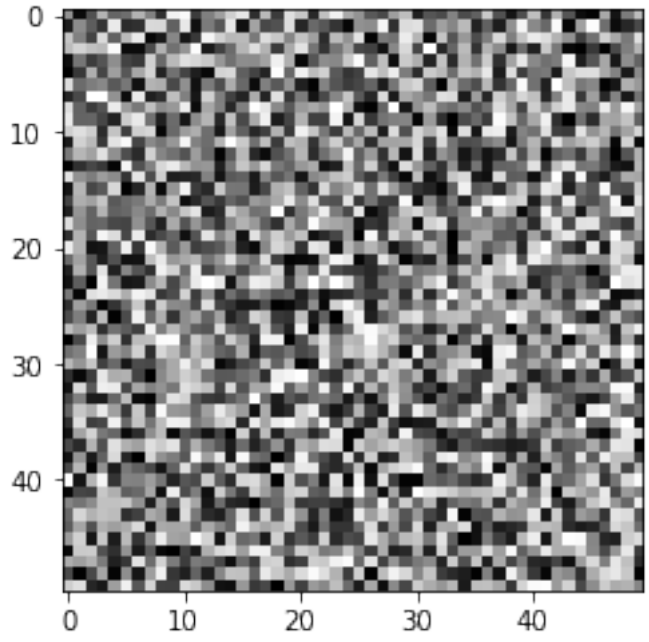
12.1 Standard Ising (on a torus)

In grayscale for fun.

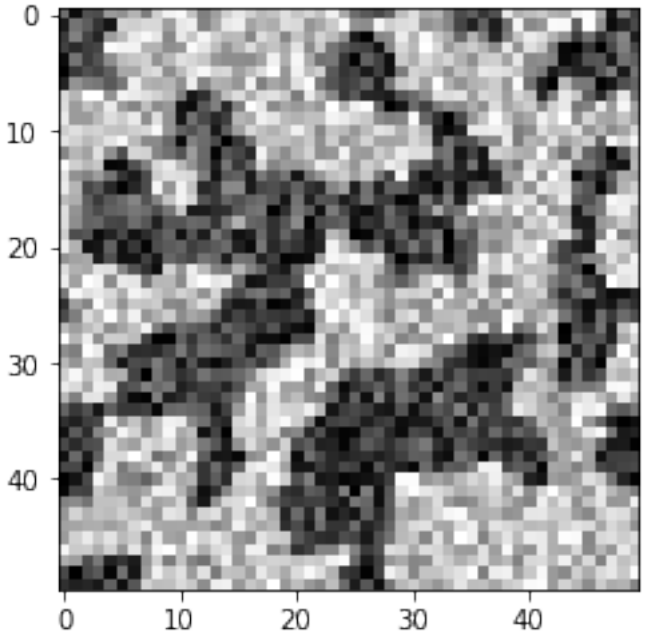
```
def neighbors(a, i, j):
    return np.hstack([a[:,j].take([i-1,i+1], mode='wrap'),
                      a[i,:].take([j-1,j+1], mode='wrap')])
```

```
def energy(img, i, j):
    return -1 + np.sum(np.abs(img[i, j] - neighbors(img, i, j)))

def isingstep(β, img):
    w, h = np.shape(img)
    i = np.random.randint(w)
    j = np.random.randint(h)
    E0 = energy(img, i, j)
    img[i, j] *= -1
    E1 = energy(img, i, j)
    P = np.exp(-β*(E1 - E0)) if E1 > E0 else 1
    if np.random.rand() > P: # Restore old
        img[i, j] *= -1
    return img
```

```
img = 2*np.random.rand(50, 50) - 1
plt.imshow(img);
```



```
n = 100000
for i in range(n):
    isingstep(3 * (np.pi / 2) / np.arctan(n - i), img)
plt.imshow(img);
```



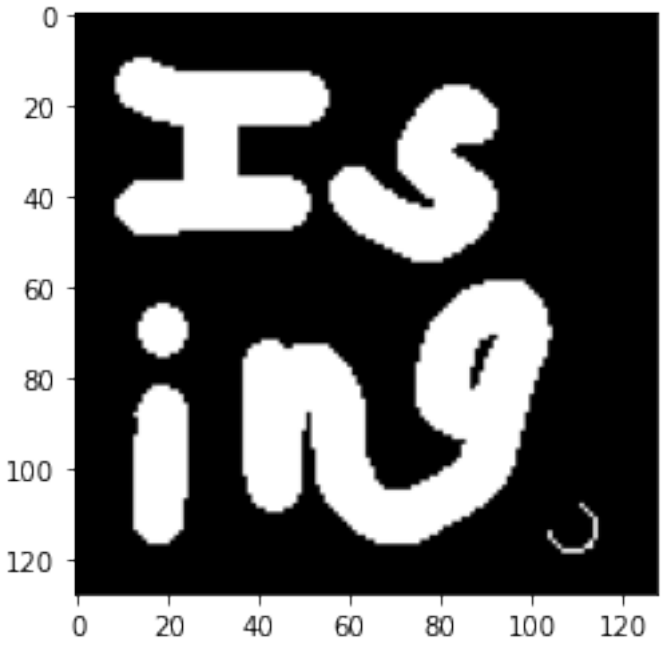
12.2 Image-edge Ising

```
edges = Image.open("ising-edges.png")
edata = np.array(edges) > 128
edges
```

```
def eenergy(img, edges, i, j):
    """Edge-modified Ising energy: 0 on edge."""
    if edges[i, j]:
        return 0
    w, h = np.shape(img)
    c = img[i, j]
    l = img[i-1, j] if i > 0   else img[w-1, j]
    r = img[i+1, j] if i < w-1 else img[0, j]
    t = img[i, j-1] if j > 0   else img[i, h-1]
    b = img[i, j+1] if j < h-1 else img[i, 0]
    return -img[i, j] * (l + r + t + b)

def nenergy(img, edges, i, j):
    """Neighbor-modified Ising energy: 0 interactions with edges."""
    if edges[i, j]:
        return 0

    w, h = np.shape(img)
    c = img[i, j]
    l = r = t = b = 0
    if i > 0:
        l = img[i-1, j] if not edges[i-1, j] else 0
    else:
        l = img[w-1, j] if not edges[w-1, j] else 0

    if i < w - 1:
        r = img[i+1, j] if not edges[i+1, j] else 0
    else:
        r = img[0, j] if not edges[0, j] else 0

    if j > 0:
        t = img[i, j-1] if not edges[i, j-1] else 0
    else:
        t = img[i, h-1] if not edges[i, h-1] else 0

    if j < h - 1:
        b = img[i, j+1] if not edges[i, j+1] else 0
    else:
        b = img[i, 0] if not edges[i, 0] else 0

    return -img[i, j] * (l + r + t + b)

def eisingstep(β, img, edges):
    w, h = np.shape(img)
    i = np.random.randint(w)
```

```
        j = np.random.randint(h)
        E0 = nenergy(img, edges, i, j)
        img[i, j] *= -1
        E1 = nenergy(img, edges, i, j)
        P = np.exp(-β*(E1 - E0)) if E1 > E0 else 1
        if np.random.rand() > P: # Restore old
            img[i, j] *= -1
        return img

def frame(writer, data):
    writer.append_data((255 * ((eimg + 1) / 2)).astype('uint8'))
```

```
img = Image.open("ising-letters.png")
eimg = -1 + 2 * (np.array(img) / 255)
plt.imshow(eimg);
```

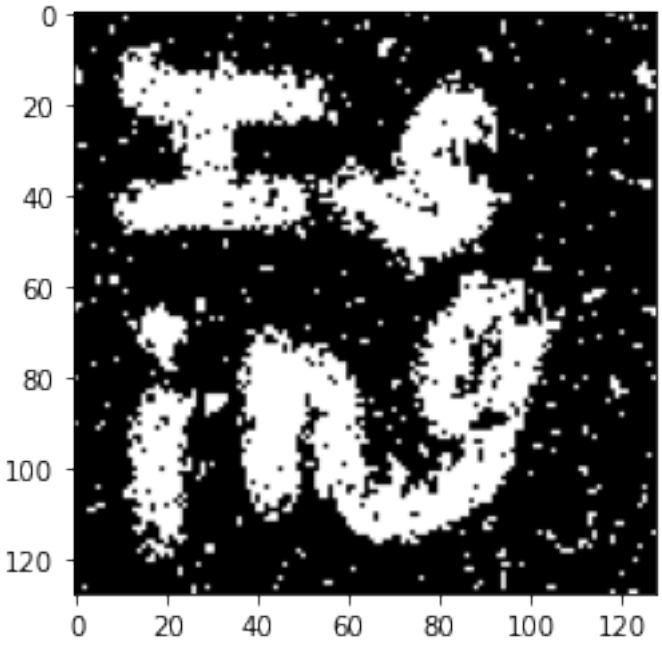


movie.gif: Full neighbor Ising.

```
n = 1000000
f = IntProgress(min=0, max=1 + (n-1) // 1000) # instantiate the bar
display(f)
with imageio.get_writer('movie.gif', mode='I') as writer:
    frame(writer, eimg)
    for i in range(n):
```

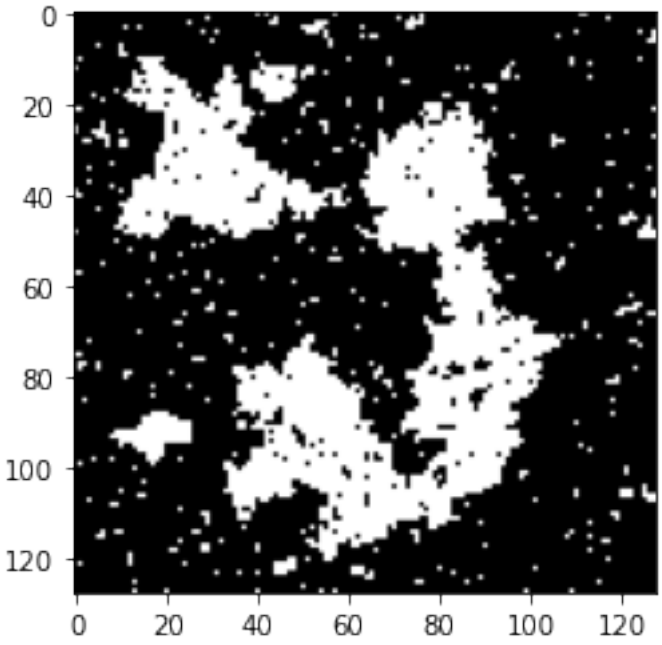
```
        eisingstep(0.5 * (np.pi / 2) / np.arctan(n - i), eimg, edata)
        if i % 1000 == 0:
            f.value += 1
            frame(writer, eimg)
plt.imshow(eimg);
```

```
IntProgress(value=0, max=1000)
```



imovie.gif: Normal Ising.

```
n = 1000000
img = eimg
with imageio.get_writer('imovie.gif', mode='I') as writer:
    frame(writer, img)
    for i in range(n):
        isingstep(0.5 * (np.pi / 2) / np.arctan(n - i), img)
        if i % 1000 == 0:
            frame(writer, img)
plt.imshow(img);
```



12.3 Image-metric Ising

```
# def takewrap(a, i, j, xs=np.arange(-1, 2), ys=np.arange(-1, 2)):
def takewrap(a, i, j, xs=np.arange(0, 1), ys=np.arange(0, 1)):
    return np.array([x for v in a.take(xs+i, axis=0, mode='wrap')
                       for x in v.take(ys+j, mode='wrap')])
```

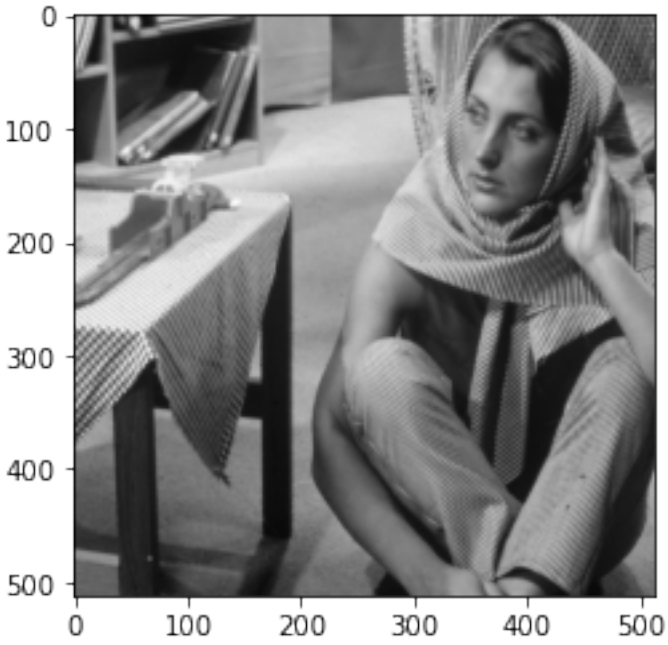
12.3.1 Unrestricted swapping motion

Swapping preserves the intensity distribution.

```
def sienergy(img, init, i, j):
    """Inversion-symmetric image energy"""
    eq = takewrap(img, i, j) == takewrap(init, i, j)
    return -np.abs(np.sum(2*eq - 1))

def ienergy(img, init, i, j):
    """Image energy based on 3x3 block deviation"""
    return np.abs(init[i, j] - img[i, j])

def swisingstep(β, img, edges):
    w, h = np.shape(img)
```

```
    i0 = np.random.randint(w)
    i1 = np.random.randint(w)
    j0 = np.random.randint(h)
    j1 = np.random.randint(h)
    E0 = ienergy(img, edges, i0, j0) + ienergy(img, edges, i1, j1)
    img[i0, j0], img[i1, j1] = img[i1, j1], img[i0, j0]
    E1 = ienergy(img, edges, i0, j0) + ienergy(img, edges, i1, j1)
    P = np.exp(-β*(E1 - E0)) if E1 > E0 else 1
    if np.random.rand() > P: # Restore old
        img[i0, j0], img[i1, j1] = img[i1, j1], img[i0, j0]
    return img

def nnisingstep(β, img, edges):
    w, h = np.shape(img)
    i0 = np.random.randint(w)
    i1 = int((i0 + np.sign(np.random.rand() - 1/2)) % w)
    j0 = np.random.randint(h)
    j1 = int((j0 + np.sign(np.random.rand() - 1/2)) % h)
    E0 = ienergy(img, edges, i0, j0) + ienergy(img, edges, i1, j1)
    img[i0, j0], img[i1, j1] = img[i1, j1], img[i0, j0]
    E1 = ienergy(img, edges, i0, j0) + ienergy(img, edges, i1, j1)
    P = np.exp(-β*(E1 - E0)) if E1 > E0 else 1
    if np.random.rand() > P: # Restore old
        img[i0, j0], img[i1, j1] = img[i1, j1], img[i0, j0]
    return img
```

```
img = Image.open("barbara.png")
eimg = -1 + 2 * (np.array(img) / 255)
initimg = eimg.copy()
plt.imshow(initimg);
```

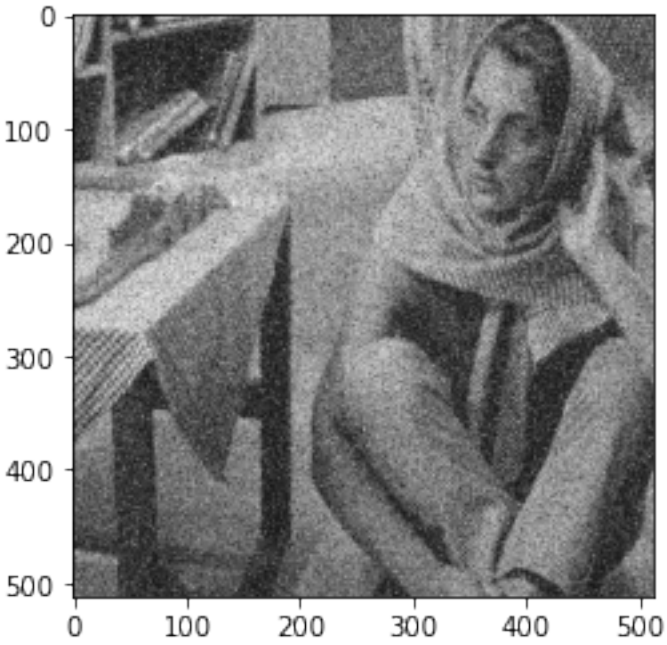


swmovie.gif: Image metric Ising (arbitrary swaps with ienergy).

```
n = 2000000
f = IntProgress(min=0, max=(1 + (n-1) // 1000)) # instantiate the bar
display(f)
with imageio.get_writer('swmovie.gif', mode='I') as writer:
    frame(writer, eimg)
    for i in range(n):
        k = i/n
        swisingstep(3, eimg, initimg)
        if i % 1000 == 0:
            f.value += 1
            frame(writer, eimg)
#     for i in range(n):
#         k = i/n
#         swisingstep(4*(1 - k) + 1e-3*k, eimg, initimg)
#         if i % 1000 == 0:
#             f.value += 1
#             frame(writer, eimg)
#     for i in range(n):
#         k = i/n
#         swisingstep(1e-3*(1 - k) + 4*k, eimg, initimg)
#         if i % 1000 == 0:
#             f.value += 1
```

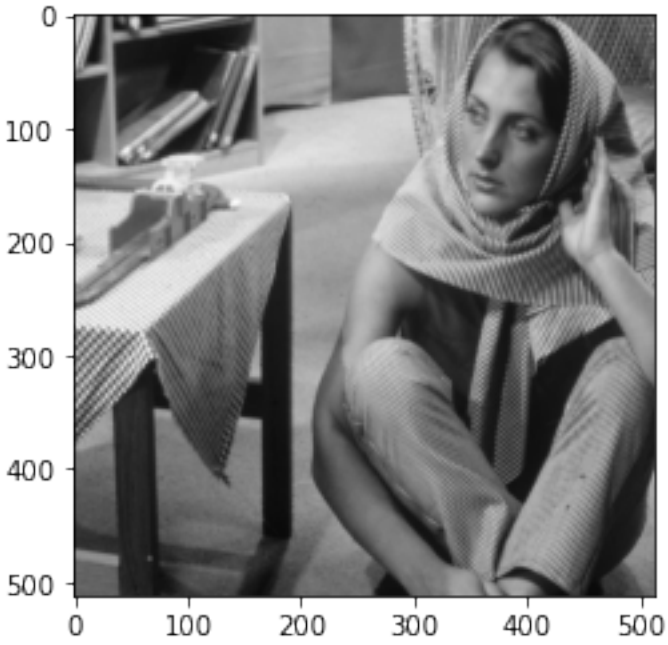
```
#             frame(writer, eimg)

plt.imshow(eimg);
```

```
IntProgress(value=0, max=2000)
```



12.3.2 Nearest-neighbor swapping motion

```
img = Image.open("barbara.png")
eimg = -1 + 2 * (np.array(img) / 255)
initimg = eimg.copy()
plt.imshow(initimg);
```

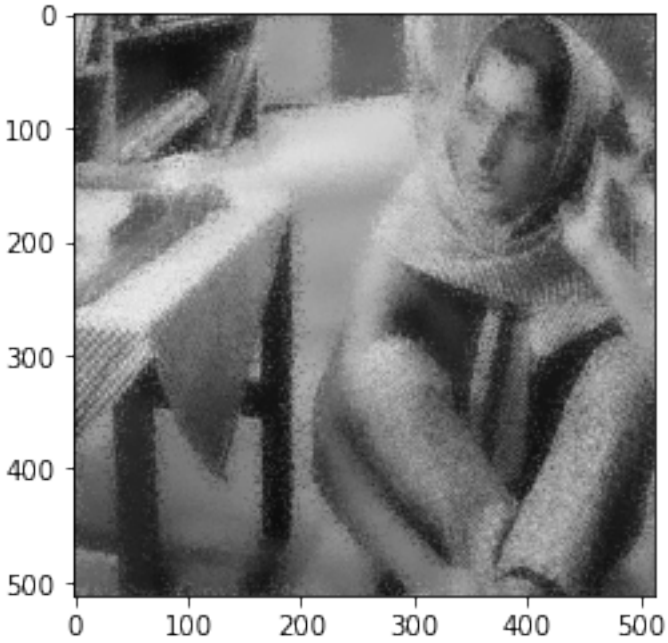


nnmovie.gif: Image metric Ising (neighborly swaps with ienergy).

```
n = 2000000
f = IntProgress(min=0, max=3*(1 + (n-1) // 1000)) # instantiate the bar
display(f)
with imageio.get_writer('nnmovie.gif', mode='I') as writer:
    frame(writer, eimg)
    for i in range(n):
        k = i/n
        nnisingstep(5*(1 - k) + 1e-4*k, eimg, initimg)
        if i % 1000 == 0:
            f.value += 1
            frame(writer, eimg)
    for i in range(n):
        nnisingstep(1e-4, eimg, initimg)
        if i % 1000 == 0:
            f.value += 1
            frame(writer, eimg)
    for i in range(n):
        k = i/n
        nnisingstep(1e-4*(1 - k) + 5*k, eimg, initimg)
        if i % 1000 == 0:
            f.value += 1
            frame(writer, eimg)
```

```
23
24 plt.imshow(eimg);
```

```
IntProgress(value=0, max=6000)
```



13 Statistical Mechanics of Images

June 3, 2020

Given the qualitative success of the image-metric based Ising images, we consider generalizations.

Definition 8. An *N-element image space* over metric spaces (K, d) and (P, a) is the space $\mathsf{Img} = (P \times K)^N$. A corresponding *image* is an element of Img.

The space P determines the spatial arrangement of the image, and is usually two-dimensional Euclidean space. We usually consider the subset of an image space where the P-coordinates are fixed, in a grid layout. The space K determines the qualities of an image at a point in P. This is usually a color or intensity space, and in practical applications is a machine integer like $128 \in \mathbb{Z}_{256}$.

Definition 9. An *image system* on an image space Img consists of a *ground image* $I_0 \in \mathsf{Img}$ and a *dispersion relation* $E : \mathbb{R} \to \mathbb{R}$. This defines the *energy* of an image

I as

$$E(I) = \sum_{(p_0,k_0)\in I_0} \sum_{k\in\{k:(p_0,k)\in I\}} E(d(k_0,\, k)).$$

Example 3. For usual images in $((\mathbb{Z}_n \times \mathbb{Z}_m) \times \mathbb{Z}_{256})^N$, where $N = nm$ and the positions of I_0 and I coincide (indexed by i and j), we have

$$E(I) = \sum_{i=1}^{n}\sum_{j=1}^{m} E(d(k_0^{ij},\, k^{ij})) = \sum_{i=1}^{n}\sum_{j=1}^{m} \varepsilon\left|k_0^{ij} - k^{ij}\right|^1,$$

with typical choices of E and d.

In the binary case ($K = \mathbb{Z}_2$), we have N independent two-level systems.

Example 4 (Grayscale images). Consider a pixel of a ground grayscale image, with integer value $k_0 \in 0, \ldots, K-1$ for even K. There are then

$$2g = 2\begin{cases} k_0, & k_0 < K/2 \\ K - k_0 - 1, & \text{else} \end{cases}$$

energy values that occur twice, and $K - 2g$ energy values that occur once (like $|x|$ on an interval like $[-3, 8]$). Thus the partition function for this single pixel is

$$\begin{aligned} Z_g &= \sum_{k=-g}^{K-g-1} e^{-\beta\varepsilon|k|} = 1 + \sum_{k=1}^{g} e^{-\beta\varepsilon k} + \sum_{k=1}^{K-g-1} e^{-\beta\varepsilon k} \\ &= 1 + \frac{e^{-\beta g\varepsilon}\left(e^{\beta g\varepsilon} - 1\right)}{e^{\beta\varepsilon} - 1} + \frac{e^{-\beta(K-g-1)\varepsilon}\left(e^{\beta(K-g-1)\varepsilon} - 1\right)}{e^{\beta\varepsilon} - 1} \end{aligned}$$

and the partition function for the whole image is

$$Z = \prod_{g=0}^{-1+K/2} Z_g^{NP(g)},$$

where $NP(g)$ is the number of pixels in the ground image with the given g-value. We then see that

$$\ln Z = \sum_{g=0}^{-1+K/2} NP(g) \ln Z_g = N \left\langle \ln Z_g \right\rangle_G,$$

where G is the random variable that takes the value g with probability $P(g)$. It then follows that $\langle E/N \rangle = \left\langle E_g \right\rangle_G$ and $S/N = \left\langle S_g \right\rangle_G$ as usual for extensive variables.

14 Thermodynamic quantities for images from a microscopic model

June 4, 2020

Since we are thinking of images as statistical entities, what is the corresponding microscopic model? Given such a model, what quantities do we consider in thermal equilibrium, and how can we understand different ensembles?

14.1 Quantum filled-site model (FSM)

Definition 10 (FSM). We define a lattice model corresponding to a *ground image* I_0 as follows. Each pixel with value $k_0 \in K \subseteq \mathbb{Z}$ in the image corresponds to a *site*, which is a discrete system with K *levels*. The energy of level k is $E_{k_0}(k) = \varepsilon|k - k_0|^r$.[1] We usually have $r = 1$ or 2. We suppose that the levels are filled by fermions that interact according to the Hamiltonian

$$\mathsf{H} = \sum_{k\in K}\sum_{i} V \mathsf{c}_{ik}^\dagger \mathsf{c}_{ik} - \sum_{\ell\in\mathcal{N}_k}\sum_{j\in\mathcal{N}_i} t_\ell \mathsf{c}_{ik}^\dagger \mathsf{c}_{j\ell}.$$

For $K \subseteq \mathbb{Z}$, $\mathcal{N}_k = k + \{-1, 0, 1\}$.

14.2 Observables and thermodynamic state variables

Several observables of the FSM are of interest:

- **Pixel colors**. The occupations n_k of different levels at a *single* site induce a distribution on K. In equilibrium, the mean level

  $$\langle k \rangle \equiv \frac{\sum_{k\in K} k n_k}{\sum_{k\in K} n_k}$$

  should be near k_0, since the energy of a level is symmetric about k_0. As the temperature increases, so will the variance of the mean level. On the flip side, does *varying* k_0 for many pixels quasistatically (changing the ground image) do work on the system? Yes, but is this consistent with what we expect?

- **Color distribution**. The net occupations m_k of different levels across *all* sites induce a distribution on K. In the special case of gray images (so levels are intensity), the entropy of the induced random variable is intensity

[1] If we want to consider colors, then K is a metric space and we replace $k - k_0$ with the metric.

entropy that we have studied previously. This is distribution is stationary when different levels cannot interact, but is it so at finite temperature?

- **Opacity**. The net occupancy of a *site* could be connected to its opacity. In equilibrium, this should be similar across all sites. Then regions with *no* particles during nonequilibrium processes make sense. The picture of a gas with fluctuating density that emits light comes to mind. When at maximum opacity, the gas in that region cannot be compressed further, and cannot accept more particles. The canonical density properties of a photon gas (like energy density) might be a good reason to choose the particles to be bosons.

- **Number of pixels**. We could vary the total number of pixels different ways. One way is to have a continuous ground image, and choose different grid discretizations. Another way is to have a large ground image grid and vary the zoom level. It would be sensible to combine these sorts of transformations with pixel color transformations, since they include translation and rotation as special cases. This seems most most similar to varying the volume of a gas. Including opacity makes fast adiabatic piston motion volume changes like $V \mapsto 2V$ make sense.

- **Number of particles**. Depending on if we allow interactions between levels, it may be appropriate to consider chemical potentials. Either for all particles, or for each color. Could this be conjugate to opacity?

- **The usual**. Given that quasistatic transformation of the other quantities does work the way we expect, we can consider the usual response variables like heat capacities and compressibilities. There is also the thermodynamic entropy.

15 Progress summary (from beginning)

June 6, 2020

Over the last two weeks, I calculated some metrics on images and did some basic simulations. The most important metric was the intensity entropy, which is used as the "entropy" of an image in the maximum entropy method (MEM) of image reconstruction used by astronomers. This was calculated for a whole image and locally in different regions of an image. On the topic of scaling, fractal dimensions were explored. The box-counting dimension was computed for different

images, and the information dimension of the intensity distribution was considered. I also did readings on probability theory and machine learning, since the usual frequentist approach that experimental physicists take is not applicable. Variants of Ising models were simulated for images, which led to the postulation of a microscopic model for varying images (the FSM), which is similar to a Hubbard model. The implications of this approach remain to be explored.

16 Progress summary

June 12, 2020

This week, I investigated different methods for obtaining thermodynamic quantities from simulations. The issue is that quantities like entropy and the Helmholtz free energy depend on global properties of the phase space (the probability or density of a microstate), and thus cannot be constructed as cumulants of microstates during a simulation. A related issue is the improbability of "tunneling" across energy barriers when taking the usual temperature-weighted steps.

These considerations motivate histogram-based methods, like the Wang-Landau algorithm that I implemented. Instead of operating in the canonical ensemble, we take a biased random walk on energies so that the result is a flat histogram of visited energies. The density of states from this process may then be used to compute a canonical partition function. Modifications where we keep a joint density of states with respect to another variable make other ensembles accessible.

Further progress with the Wang-Landau algorithm was slowed by a discrepancy in the steps needed between Wang and Landau's results and mine for the 32 by 32 Ising ferromagnet. Their original paper claims a 0.035 % average error in the density of states after only 700 000 total spin flips.[2] This is far fewer than the spin flips I needed, and another paper that looks at the scaling of the tunneling time (spin flips to go from ground to anti-ground state) might corroborate this.[3] Their tunneling times are all well above 1 720 000 (an eighth of their τ_{exact}) for the same simulation. Success in the Wang-Landau algorithm requires visiting all energies many times, so execution takes several tunneling times. We are still trying to resolve this discrepancy, as well as other vague details from the original paper, like the possible choice of energy bins for continuous systems and unspecified edge behavior for energy intervals.

I have looked at other histogram methods like WHAM, as well as parallel tempering (replica exchange MCMC). Both parallel tempering and the Wang-Landau

[2] 10.1103/PhysRevLett.86.2050

[3] 10.1103/PhysRevLett.92.097201

algorithm are attractive because they are easily parallelizable. I have also come back to considering the MAXENT/MEM image reconstruction technique and the issues of entropy and feature representation again.

17 Description of MaxEnt

June 13, 2020

I describe the basic idea of the MaxEnt reconstruction technique, and implement the most naïve approach. This makes it clear that a better optimization algorithm is needed. The idea here is less to make it work than to get an idea of how it works, and then use a better software package later on if we do end up using MaxEnt.

18 Maximum-entropy reconstruction

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from scipy import signal, misc
```

Given a measurement D of a system I_0, we wish to reconstruct I_0 from D. We take the Bayesian approach and find the I that maximizes $P(I|D) \propto P(D|I)P(I)$, where the likelihood is related to the error in D due to I and the prior is $P(I) = \exp(\lambda S(I))$, where S is some notion of the entropy of an image. If $D = T(I_0)$, then $P(D|I) = \exp(-E(T(I), D))$, for some metric E on D's. From the form of the objective $\ln P(D|I) + \ln(I) = -E(T(I), D) + \lambda S(I)$, we see that the entropy acts as a regularizer.

The astronomers have both I and D as intensity lists $\mathbb{Z}_N \to \mathbb{R}_{\geq 0}$, where

$$S(I) = \sum_i I_i \ln I_i$$

and

$$E(D', D) = \sum_i \exp\left(-\frac{(D'_i - D_i)^2}{2\sigma_i}\right)$$

(with empirical σ_i). The transformation T is convolution with the point spread function of the telescope.

To perform the optimization, we need not only the functions S and E, but also a procedure to modify candidate images.

```
def maxent_objective(D, I, λ): # For minimization
    return D.E(I.transform()) - λ*I.S

# Greedy for now, but can be something like simulated annealing
def maxent(D, I, λ = 1, N = 1_000_000, ε = 1e-8):
    f0 = np.inf
    f = maxent_objective(D, I, λ)
    i = 0
    while ε < f0 - f and i < N:
        I.propose()
        fv = maxent_objective(D, I, λ)
        if fv < f: # Greedy
            f0, f = f, fv
            I.accept()
        i += 1
        if i % (N // 20) == 0:
            print("Maxent: {} / {}".format(i, N))

    print("Maxent: i: {} / {}, Δf: {}".format(i, N, f0 - f))
    return I, i, f
```

18.1 Example: 1D Point from Gaussian

```
class Gaussian:
    def __init__(self, μ=0, σ=1):
        self.μ = μ
        self.σ = σ
    def E(self, Gv):
        return (self.μ - Gv.μ)**2 + (self.σ - Gv.σ)**2

class Point:
    def __init__(self, x=0):
        self.x = x
        self.dx = 0
        self.S = self.entropy()
    def propose(self):
        self.dx = np.random.rand() - 0.5
        self.S = self.entropy()
    def accept(self):
        self.x += self.dx
        self.dx = 0 # Idempotence
    def entropy(self):
        return -(self.x - 10)**2 # Opposite from true max
    def transform(self):
        return Gaussian(μ = self.x + self.dx)
```

```
D = Gaussian(0, 1)
I = Point(9)
I0, _, _ = maxent(D, I, λ = 1, N = 1_000_000, ε = 1e-4)
I0.x
```

```
Maxent: i: 1000000 / 1000000, Δf: 0.011299906795997572
```

```
4.268883578482481
```

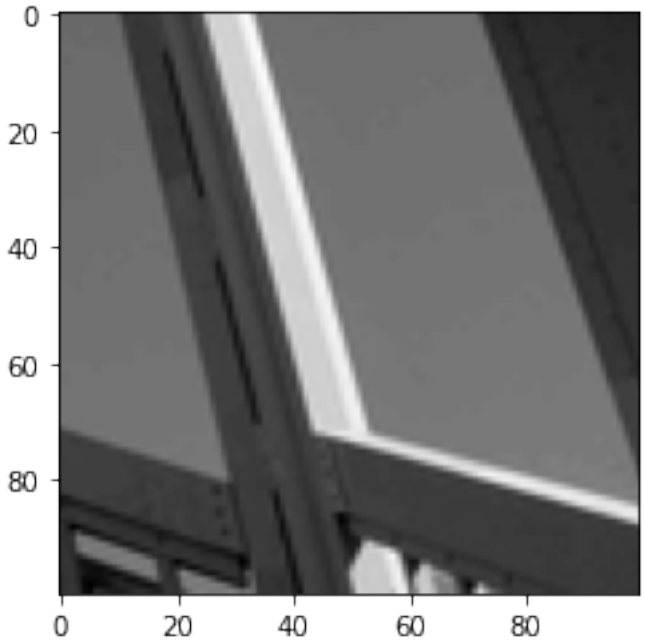
Results are sort of near 5. The optimization is terrible, but you get the idea: the entropy shifts the best point away from zero.

18.2 Example: Image from PSF convolution (measurement)

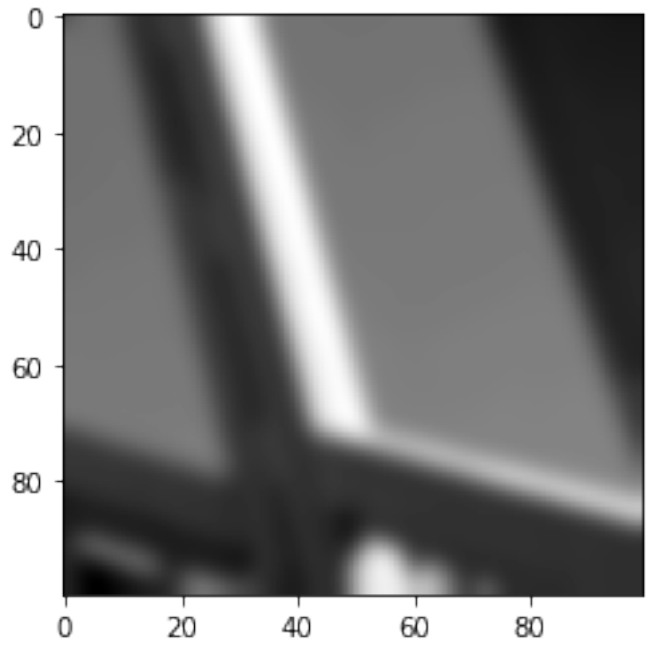
```
class DImage:
    def __init__(self, I):
        self.I = I
    def E(self, Dv):
        return np.sum((self.I - Dv.I)**2)

class IImage:
    def __init__(self, I):
        self.I = I
        self.w, self.h = np.shape(I)
        self.i, self.j = 0, 0
        self.I0 = 0
        self.Iv = 0
        n = int(np.sqrt(self.w * self.h))
        self.G = signal.windows.gaussian(n // 10, n // 50)
        self.S = self.entropy()
    def propose(self):
        self.i, self.j = np.random.randint(self.w), np.random.randint(self.h)
        I0 = self.I[self.i, self.j]
        self.I0 = I0
        self.Iv = np.random.randint(256)
        self.S = self.entropy()
    def accept(self):
        self.I[self.i, self.j] = self.Iv
    def entropy(self):
```

```
        return -np.sum(np.log(self.I + 1)) + self.I0*np.log(self.I0 + 1) -
        ↪  self.Iv*np.log(self.Iv + 1)
    def transform(self):
        Iv = self.I.copy()
        Iv[self.i, self.j] = self.Iv
        return DImage(signal.sepfir2d(Iv, self.G, self.G))
```

```
I = IImage(misc.ascent()[250:350,250:350])
plt.imshow(I.I, cmap='gray');
```




```
plt.imshow(I.transform().I, cmap='gray');
```

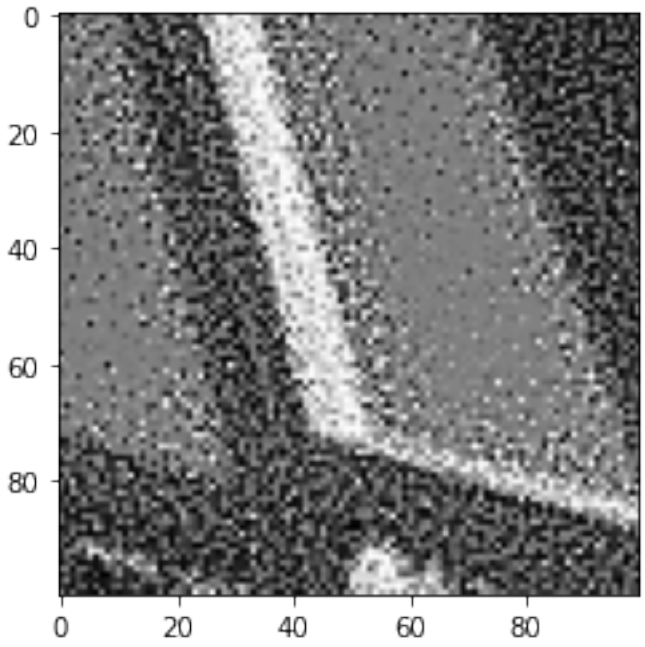



```
I = IImage(misc.ascent()[250:350,250:350])
Iguess = IImage(128 * np.ones((I.w, I.h), dtype=int))
I0, _, _ = maxent(I.transform(), Iguess, λ = 1e-9, N = 1_000_000, ε = 1e-20) # Just do the max iterations
```

```
Maxent: 50000 / 1000000
Maxent: 100000 / 1000000
Maxent: 150000 / 1000000
Maxent: 200000 / 1000000
Maxent: 250000 / 1000000
Maxent: 300000 / 1000000
Maxent: 350000 / 1000000
Maxent: 400000 / 1000000
Maxent: 450000 / 1000000
Maxent: 500000 / 1000000
Maxent: 550000 / 1000000
Maxent: 600000 / 1000000
Maxent: 650000 / 1000000
Maxent: 700000 / 1000000
Maxent: 750000 / 1000000
Maxent: 800000 / 1000000
```

```
Maxent: 850000 / 1000000
Maxent: 900000 / 1000000
Maxent: 950000 / 1000000
Maxent: 1000000 / 1000000
Maxent: i: 1000000 / 1000000, Δf: 4.847227362683043
```

```
plt.imshow(I0.I, cmap='gray');
```



19 “Greedy” painting-like pictures

June 14, 2020

We noticed that swapping neighboring pixels in the previous simulations produced images that looked somewhat like paintings. There are other ways to make images look like paintings, and this is one of them. We randomly draw a shape of fixed size on the image. The color of the shape is the mean color of the image in the rectangle where the shape is drawn. This is done many times, and the size of the shapes gradually decreases. I call this “greedy” because choosing the mean color is the locally optimal color. This method has the advantage of being similar to the process of painting, which is why it is presented instead of the usual “oilify” filters in image processing software. An improvement would try to

draw shapes that reflect the shapes of the subject matter, and perhaps add other texture. These kinds of considerations might require more global strategies.

20 Greedy Cubism

Draw an image by greedily drawing cubes.

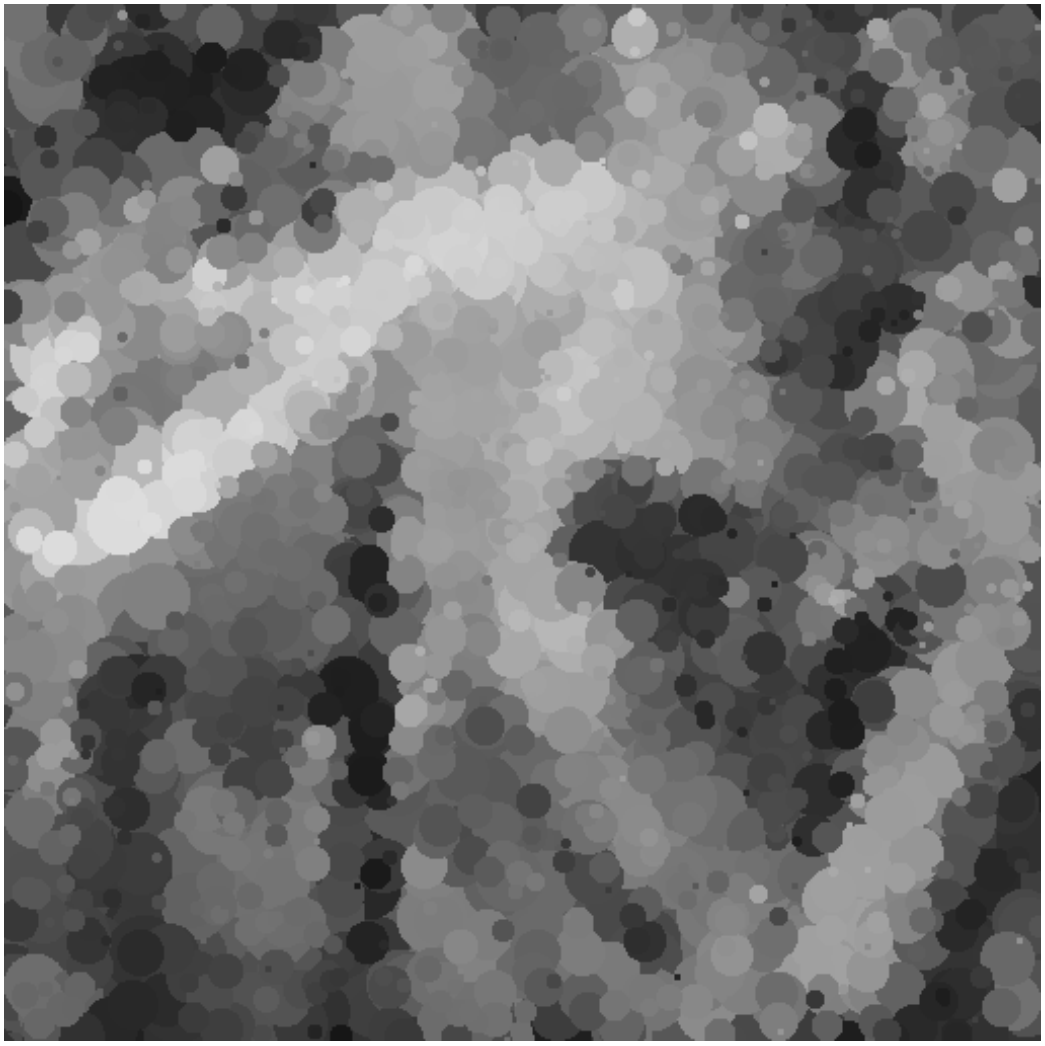
```
import numpy as np
import matplotlib.pyplot as plt
from PIL import Image, ImageOps, ImageDraw
from scipy import signal, misc
```

```
img = Image.open("barbara.png")
```

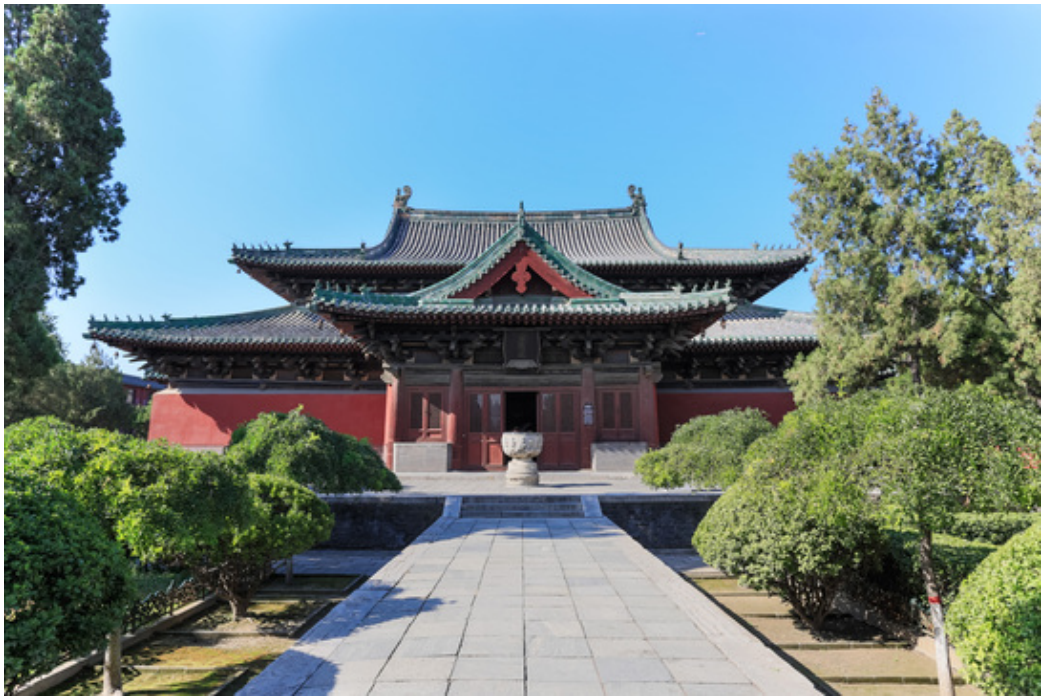
```
def greedy_cubes(img, N):
    xs, ys = img.size
    art = Image.new(img.mode, (xs, ys))
    draw = ImageDraw.Draw(art)
    rmax = int(np.sqrt(xs*ys) / 10)
    r = rmax
    ε = 10
    for i in range(N):
        x = np.random.randint(xs)
        y = np.random.randint(ys)
        [np.mean(c) for c in cimg.split()]
        r = int(rmax * (1 - (i/(N+1))**2)) + 1
        box = [x - r, y - r, x + r, y + r]
        color = tuple(int(np.round(np.mean(c))) for c in img.crop(box=box).split())
        draw.ellipse(box, fill=color)
    return art
```

```
art = greedy_cubes(img, 10000)
```

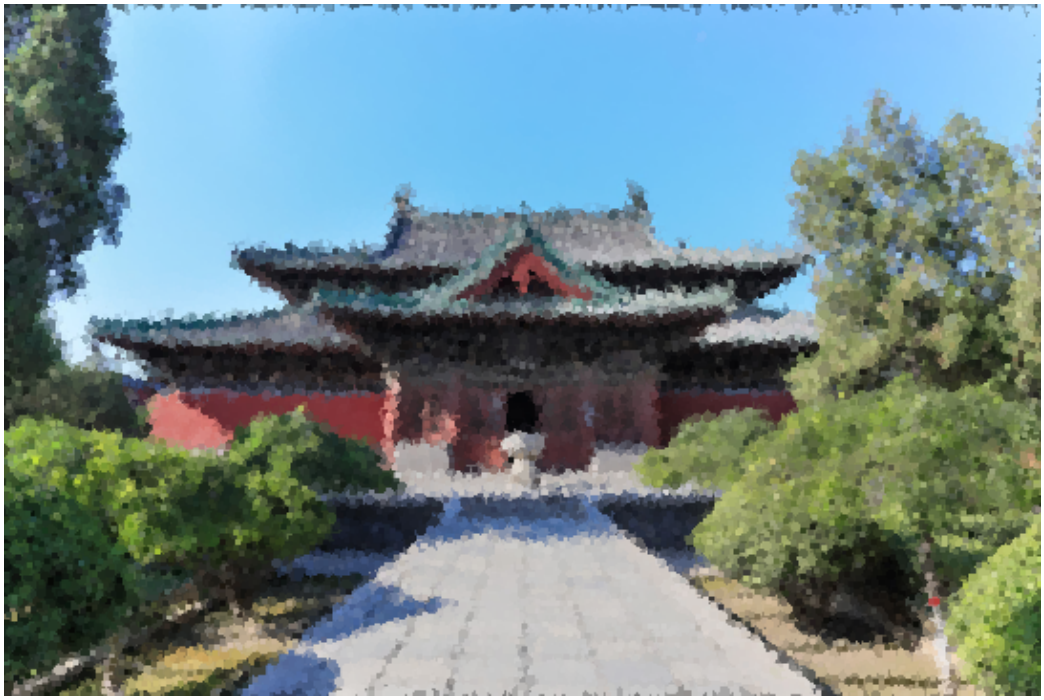
```
art
```



```
cimg = Image.open('test.jpg')
cimg
```



```
art = greedy_cubes(cimg, 1000000)
art
```



21 Exact density of states for gray and BW images

June 16, 2020

In doing the future statistical simulations of images, we would like to know the density of states $g(E)$ in our site-based model. This is the number of ways to choose the N pixel values m_n from $0, \ldots, M$ so that

$$E = \sum_{n=1}^{N} \left| m_n - m_n^* \right|,$$

where $0 \leq E \leq MN$ and m_n^* is the value of the ground image at pixel n. While the solution to follow works for arbitrary values of m_n^*, the results are simplest (no multinomial coefficients) if we consider $m_n^* = 0$ or M or $m_n^* = M/2$. That is, the image is all black or white (BW) or gray. Then we have

$$E = \sum_{n=1}^{N} g' m_n$$

with $m_n = 0, \ldots, M$ and $g' = 1$ (BW) or $m_n = 0, \ldots, M/2$ and $g' = 2$ (gray).[4] In that case, we'll consider BW for concreteness.

We encode the energy as an exponent, so that $g(E)$ is the coefficient of x^E in the polynomial

$$p(x) \equiv \left(x^0 + x^1 + \cdots + x^M\right)^N,$$

since each inner sum represents different assignments of a pixel value and the products will produce terms that represent all different combinations of values. We then expand to find

$$\begin{aligned} p(x) &= \left(\frac{1 - x^{M+1}}{1 - x}\right)^N \\ &= \sum_{k=0}^{N} \sum_{j=0}^{\infty} (-1)^k \binom{N}{k} x^{(M+1)k} \cdot (-1)^j \binom{-N}{j} x^j \\ &= \sum_{k=0}^{N} \sum_{j=0}^{\infty} (-1)^k \binom{N}{k} \binom{N+j-1}{j} x^{(M+1)k+j}. \end{aligned}$$

The value of j in the inner sum for x to have power E is $j = E - k(M+1)$, so the coefficient of x^E in $p(x)$ is

$$g(E) = \sum_{k=0}^{N} (-1)^k \binom{N}{k} \binom{N + E - k(M+1) - 1}{E - k(M+1)}.$$

(Due to the stucture of the binomial coefficients, we may let the summation extend over all $k \in \mathbb{Z}$ for further manipulation.)

22 Simulations for canonical ensemble averages

June 18, 2020

The most promising seem to be:

- Wang-Landau
- Parallel tempering (replica exchange MCMC).

[4] This gives $m = 0$ degeneracy 2, which is not strictly correct but matters little.

23 Systems for organized simulation

```
import sys
if 'src' not in sys.path: sys.path.append('src')
```

List of imports for all relevant systems.

```
from statistical_image import *
```

23.1 Specification

TODO: - Encode this specification and the requirements of simulations as abstract base classes? - Consider changing to `numba.typed.Dict` in the future if the API is guaranteed to be stable.

A system for simulation is a Numba `jitclass` that implements `state`, `state_names`, and `copy` functions. Given an instance `s` of a system class `System`, these function should satisfy

```
id_systems = [
    s,
    s.copy(), # deep
    System(*s.state()),
    System(**{k: v for k, v in zip(s.state_names(), s.state())})
]
1 == len({t.state() for t in id_systems}) == len({t.state_names() for t in id_systems})
```

By default, we have

```
class System: # ...
    def copy(self):
        return self.__class__(*self.state())
```

In addition, different simulations may require more methods to be implemented. ### Wang-Landau A Wang-Landau simulation requires a `System` to have the variables `E` and `Ev` and to implement the methods `energy_bins`, `energy`, `propose`, and `accept`.

24 System: Statistical Image

```
import numpy as np
from numba.experimental import jitclass
from numba import int64
integer = int64
```

```
__all__ = ['StatisticalImage']
@jitclass([
    ('I0', integer[:]),
    ('I',  integer[:]),
    ('N',  integer),
    ('M',  integer),
    ('E',  integer),
    ('Ev', integer),
    ('dE', integer),
    ('dx', integer),
    ('i',  integer)
])
class StatisticalImage:
    def __init__(self, I0, I, M):
        if len(I0) != len(I):
            raise ValueError('Ground image I0 and current image I should have the same length.')
        if M < 0:
            raise ValueError('Maximum site value must be nonnegative.')
        self.I0 = I0
        self.I  = I
        self.N  = len(I0)
        self.M  = M
        self.E  = self.energy()
        self.Ev = self.E
        self.dE = 0
        self.dx = 0
        self.i  = 0
    def state(self):
        return self.I0.copy(), self.I.copy(), self.M
    def state_names(self):
        return 'I0', 'I', 'M'
    def copy(self):
        return StatisticalImage(*self.state())
    def energy_bins(self):
        E0 = 0
        Ef = np.sum(np.maximum(self.I0, self.M - self.I0))
        ΔE = 1
        return np.arange(E0, Ef + ΔE + 1, ΔE)
    def energy(self):
```

```
        return np.sum(np.abs(self.I - self.I0))
    def propose(self):
        i = np.random.randint(self.N)
        self.i = i
        x0 = self.I0[i]
        x = self.I[i]
        r = np.random.randint(2)
        if x == 0:
            dx = r
        elif x == self.M:
            dx = -r
        else:
            dx = 2*r - 1
        dE = np.abs(dx) if x0 == x else (dx if x0 < x else -dx)
        self.dx = dx
        self.dE = dE
        self.Ev = self.E + dE
    def accept(self):
        self.I[self.i] += self.dx
        self.E = self.Ev
```

25 Organized parallel simulations

```
import numpy as np
from multiprocessing import Pool
from scipy.signal import windows
import sys
import time
import os, struct # for `urandom`
import pprint # for parameters
import tempfile
import h5py, hickle
```

Since the Numba `jitclass` objects are not picklable, the relevant parameters to reconstruct a system are passed between processes. We require that all systems be accessible from the `systems` module.

```
if 'src' not in sys.path: sys.path.append('src')
import systems
```

```
def make_params(system):
    return {system.__class__.__name__:
            {k: v for k, v in zip(system.state_names(), system.state())}}
```

```
4
5   def make_system(system_params, system_prep = lambda x:x):
6       return system_prep([getattr(systems, cls)(**state)
7                           for cls, state in system_params.items()][0])

1   def make_psystems(params, psystem_prep): #:: params -> (system -> [system]) -> [params]
2       log = params.get('log', False)
3       if log:
4           print('Finding parallel bin systems ... ', end='', flush=True)
5       psystems = psystem_prep(make_system(params['system']), **params['parallel'])
6       if log:
7           print('done.')
8       return [(make_params(s), *r) for s, *r in psystems]

1   def urandom_reseed():
2       """Reseeds numpy's RNG from `urandom` and returns the seed."""
3       seed = struct.unpack('I', os.urandom(4))[0]
4       np.random.seed(seed)
5       return seed
6
7   def worker(simulation, psystem, params):
8       log = params.get('log', False)
9       urandom_reseed()
10      psystem_params, *args = psystem
11      system = make_system(psystem_params)
12      if log:
13          print('(', end='', flush=True)
14      # Individual simulation output is too much when running parallel simulations.
15      params['simulation'].update({'log': False})
16      results = simulation(system, *args, **params['simulation'])
17      if log:
18          print(')', end='', flush=True)
19      return results
20
21  def show_params(params):
22      print('Run parameters')
23      print('--------------')
24      pprint.pp(params, sort_dicts=False)
25      print()
26
27  def save_results(results, params, log=False, prefix='simulation-', dir='data'):
28      with tempfile.NamedTemporaryFile( # Note: dir shadows dir()
29          mode='wb', prefix=prefix, suffix='.h5', dir=dir, delete=False) as f:
30          with h5py.File(f, 'w') as hkl:
31              if log:
32                  print('Writing results ... ', end='', flush=True)
```

```
            hickle.dump({
                'parameters': params,
                'results': results
            }, hkl)
            relpath = os.path.relpath(f.name)
            if log:
                print('done: ', relpath)
    return relpath

def run(params, simulation, system_prep,
        psystem_prep = lambda x:x, result_wrapper = lambda x:x, **kwargs):
    params.update(kwargs)
    parallel = 'parallel' in params
    log = params.get('log', False)
    if log:
        show_params(params)

    if parallel:
        psystems = make_psystems(params, psystem_prep)
    else:
        psystem = make_system(params['system'], system_prep)

    if log:
        if parallel:
            print('Running || ', end='', flush=True)
        else:
            print('Running ...')
        start_time = time.time()

    if parallel:
        with Pool() as pool:
            results = pool.starmap(worker, ((simulation, args, params) for args in psystems))
        results = [result_wrapper(r) for r in results]
    else:
        results = result_wrapper(simulation(*psystem, **params['simulation'], **kwargs))

    if log:
        seconds = int(time.time() - start_time)
        if parallel:
            print(' || done in', seconds, 'seconds.')
        else:
            print('... done in', seconds, 'seconds.')

    # Save single-shot results in a singleton list so that we can analyze parallel and
    # single results the same way.
```

```
    rdict = {'results': results if parallel else [results]}
    save_params = params.pop('save', False)
    if save_params:
        relpath = save_results(results, params, log, **save_params)
        rdict.update({'file': relpath})

    return rdict
```

We can choose overlapping bins for the parallel processes to negate boundary effects.

```
def extend_bin(bins, i, k = 0.05):
    if len(bins) ≤ 2: # There is only one bin
        return bins
    k = max(0, min(1, k))
    return (bins[i] - (k*(bins[i] - bins[i-1]) if 0 < i else 0),
            bins[i+1] + (k*(bins[i+2] - bins[i+1]) if i < len(bins) - 2 else 0))
```

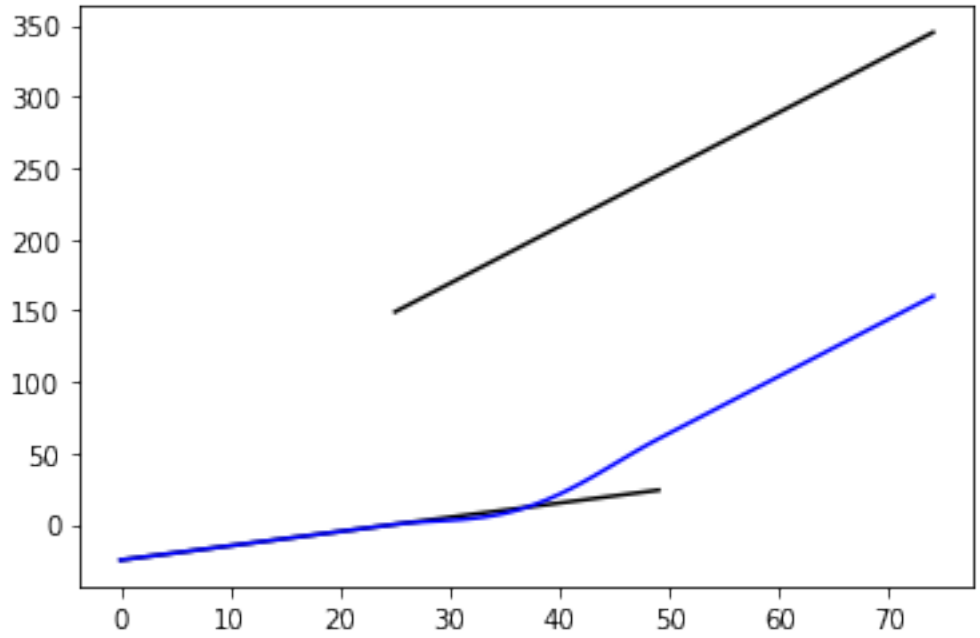
Often parallel results are the value of a real function on some grid or list of bins. Given that many of these pieces may overlap, we must combine them back together into a full solution. This requires first transforming the results so that they are comparable, and then performing the combination. The most common case is repetition of the same real-valued experiment. No transformation is required, and we simply average all the results. Even better, we may assign the values within each piece a varying credence from 0 to 1 and perform weighted sums.

```
def join_results(xs, ys, wf = windows.hann):
    xf = np.array(sorted(set().union(*xs)))
    xi = [np.intersect1d(xf, x, return_indices=True)[1] for x in xs]

    n, m = len(xf), len(xs)
    ws = np.zeros((m, n))
    wc = np.zeros((m, n))
    for i in range(m):
        l = len(xs[i])
        ws[i, xi[i]] = wf(l)
        wc[i, xi[i]] = np.ones(l)
    unweighted = np.sum(wc, 0) ≤ 1

    Δys = np.zeros(m)
    for i in range(m):
        Σc = Σw = 0
```

```
        for j in range(i):
            a = Δys[j] * np.ones(n)
            a[xi[j]] += ys[j]
            a[xi[i]] -= ys[i]
            w = ws[i,:] * ws[j,:]
            Σc += np.dot(a, w)
            Σw += np.sum(w)
        Δys[i] = Σc / Σw if i > 0 else 0

    yf = np.zeros(n)
    for i in range(m):
        w = ws[i, xi[i]]
        # The weights are meaningful only as relative weights at overlap points.
        # We must avoid division by zero at no-overlap points with weight zero.
        # Note that overlap points with all weights zero will be an issue, as
        # the weights in that situation are meaningless.
        w[(w == 0) & unweighted[xi[i]]] = 1
        yf[xi[i]] += (ys[i] + Δys[i]) * w
        ws[i, xi[i]] = w
    Σws = np.sum(ws, 0)
    yf /= Σws

    return xf, yf, Δys
```

Demonstration of joining overlapping results.

```
if __name__ == '__main__':
    from matplotlib import pyplot as plt

    testn = 50
    xs = [np.arange(testn), np.arange(testn // 2, testn + testn // 2)]
    ys = [np.arange(testn) - testn // 2, 4*np.arange(testn // 2, testn + testn // 2) + (testn -
    ↪  1)]
    axs, ays, _ = join_results(xs, ys)

    for x, y in zip(xs, ys):
        plt.plot(x, y, 'black')
    plt.plot(axs, ays, 'blue');
```



26 The Wang-Landau algorithm (density of states)

We determine thermodynamic quantities from the partition function by obtaining the density of states from a simulation.

```
from numba import njit
import numpy as np
```

```
import sys
if 'src' not in sys.path: sys.path.append('src')
import simulation as sim
```

Utility functions for the simulation.

```
@njit(inline='always')
def bisect_right(a, x, lo=0, hi=None):
    if lo < 0:
        raise ValueError('lo must be non-negative')
    if hi is None:
        hi = len(a)
    while lo < hi:
        mid = (lo + hi) // 2
```

```
        if x < a[mid]:
            hi = mid
        else:
            lo = mid + 1
    return lo

@njit
def binindex(a, x):
    return bisect_right(a, x, lo=0, hi=len(a) - 1) - 1
```

```
@njit
def flat(H, ε = 0.2):
    """Determines if a histogram is approximately flat to within ε of the mean height."""
    return not np.any(H < (1 - ε) * np.mean(H)) and np.all(H ≠ 0)
```

26.1 Algorithm

A Wang-Landau algorithm, with quantities as logarithms and with monte-carlo steps proportional to $f^{-1/2}$ (a "Zhou-Bhat schedule").

We use energy bins encoded by numbers E_i for $i \in [0,\, N]$, so that there are N bins. The energies E covered by bin i satisfy $E_i \le E < E_{i+1}$. For the bounded discrete systems that we are considering, we must choose E_N to be an arbitrary number above the maximum energy.

```
def system_prep(system):
    return system, system.energy_bins()
```

```
@njit
def simulation(system,
               Es,            # The energy bins
               M = 1_000_000, # Monte carlo step scale
               eps = 1e-8,    # f tolerance
               logf0 = 1,     # Initial log f
               flatness = 0.2, # Desired histogram flatness
               log = False    # Log progress of f-steps
              ):
    if M ≤ 0 or eps ≤ 1e-16 or not (0 < logf0 ≤ 1) or not (0 ≤ flatness < 1):
        raise ValueError('Invalid Wang-Landau parameter.')

    # Initial values
    E0 = Es[0]
    Ef = Es[-1]
    N = len(Es) - 1
    logf = 2 * logf0
```

```
    logftol = np.log(1 + eps)
    S = np.zeros(N) # Set all initial g's to 1
    H = np.zeros(N, dtype=np.int32)
    i = binindex(Es, system.E)
    converged = True
    steps = 0

    if log:
        fiter = 0
        fiters = int(np.ceil(np.log2(logf0) - np.log2(logftol)))
        print("Wang-Landau START")

    while logftol < logf:
        H[:] = 0
        logf /= 2
        iters = 0
        niters = int((M + 1) * np.exp(-logf / 2))
        if log:
            fiter += 1
        while not flat(H, flatness) and iters < niters:
            system.propose()
            Ev = system.Ev
            j = binindex(Es, Ev)
            if E0 ≤ Ev < Ef and (
                S[j] < S[i] or np.random.rand() < np.exp(S[i] - S[j])):
                system.accept()
                i = j
            H[i] += 1
            S[i] += logf
            iters += 1
        steps += iters
        if niters ≤ iters:
            converged = False
        if log:
            print("f: ", fiter, " / ", fiters, "\t(", iters, " / ", niters, ")")

    if log:
        print("Done: ", steps, " total MC iterations.")
    return Es, S, H, steps, converged

def wrap_results(results):
    return {k: v for k, v in zip(('Es', 'S', 'H', 'steps', 'converged'), results)}
```

26.1.1 Parallel construction of the density of states

```
@njit
def find_bin_systems(system, Es, Ebins, N = 1_000_000, method = 'wl'):
    """
    Find systems with energies in the bins given by `Es` by stepping `sys`.

    Args:
        system: The initial system to search from. This is usually a ground state.
        Es: The energies of the system.
        Ebins: The energy bins to find systems for.
        N: The maximum number of steps to try.
        method: The string name of the search method to try.
            'wl': Wang-Landau steps where we prefer energies we have not visited
            'increasing': Only accept increases in energy. This only works for
                steps that are not trapped by local maxima of energy.

    Returns:
        A list of independent systems with energies in Ebins.

    Raises:
        ValueError: The method argument was invalid.
        RuntimeError: Bin systems could not be found after N steps.
    """
    if method == 'wl':
        S = np.zeros(len(Es), dtype=np.int32)
    systems = [None] * (len(Ebins) - 1)
    n = 0
    l = len(Ebins) - 1
    systems = [system] * l
    empty = np.repeat(True, l)
    i = binindex(Es, system.E)
    while np.any(empty) and n < N:
        for s in range(l):
            if empty[s] and Ebins[s] ≤ system.E < Ebins[s + 1]:
                systems[s] = system.copy()
                empty[s] = False

        system.propose()
        j = binindex(Es, system.Ev)
        if method == 'wl':
            if S[j] < S[i]:
                i = j
                system.accept()
            S[i] += 1
        elif method == 'increasing':
```

```
            if system.E < system.Ev:
                system.accept()
        else:
            raise ValueError('Invalid method argument for finding bin systems.')
        n += 1

    if N ≤ n:
        raise RuntimeError('Could not find bin systems (hit step limit).')
    return systems
```

```
def psystem_prep(system, bins = 8, overlap = 0.1, steps = 1_000_000, method = 'wl', **kwargs):
    Es = system.energy_bins() # Intrinsic to the system
    Ebins = np.linspace(Es[0], Es[-1], bins + 1) # For parallel subsystems
    systems = find_bin_systems(system, Es, Ebins, steps, method)
    binEs = [(lambda E0, Ef: Es[(E0 ≤ Es) & (Es ≤ Ef)])(*sim.extend_bin(Ebins, i, overlap))
             for i in range(len(Ebins) - 1)]
    return zip(systems, binEs)
```

```
def run(params, **kwargs):
    return sim.run(params, simulation, system_prep, psystem_prep, wrap_results, **kwargs)
```

```
def join_results(results, *args, **kwargs):
    return sim.join_results(*zip(*[(r['Es'][:-1], r['S']) for r in results]), *args, **kwargs)
```

26.2 Thermal calculations on images

```
from numba import njit
from numba.experimental import jitclass
from numba import int64
integer = int64
```

```
import numpy as np
from scipy import interpolate, special
import os
import tempfile
import h5py, hickle
import pprint
```

```
import sys
if 'src' not in sys.path: sys.path.append('src')
import simulation as sim
import wanglandau as wl
```

26.2.1 Parallel Simulation

```
N = 16
Moff = 0
I0 = Moff * np.ones(N, dtype=int)
system_params = {
    'StatisticalImage': {
        'I0': I0,
        'I': I0.copy(),
        'M': 2**5 - 1
    }
}
```

```
params = {
    'system': system_params,
    'simulation': {
        'M': 1_000_000,
        'eps': 1e-8,
        'logf0': 1,
        'flatness': 0.2
    },
    'parallel': {
        'bins': 2,
        'overlap': 0.25,
        'steps': 1_000_000
    },
    'save': {
        'prefix': 'simulation-',
        'dir': 'data'
    }
}
```

```
# params.pop('parallel', None) # Single run
wlresults = wl.run(params, log=True)
```

```
Run parameters
--------------
{'system': {'StatisticalImage': {'I0': array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]),
                                 'I': array([ 0,  7, 16, 23, 18,  3, 10,  5, 17,  6,  3,  8,  2, 20, 10,
                                 'M': 31}},
 'simulation': {'M': 1000000, 'eps': 1e-08, 'logf0': 1, 'flatness': 0.2},
 'parallel': {'bins': 2, 'overlap': 0.25, 'steps': 1000000},
 'save': {'prefix': 'simulation-', 'dir': 'data'},
 'log': True}
```

```
Finding parallel bin systems ... done.
Running || (()) || done in 17 seconds.
Writing results ... done:  data/simulation-l3t1t_sn.h5
```

```
wlEs, S, ΔS = wl.join_results(wlresults['results'])
```

```
[(r['steps'], r['converged']) for r in wlresults['results']]
```

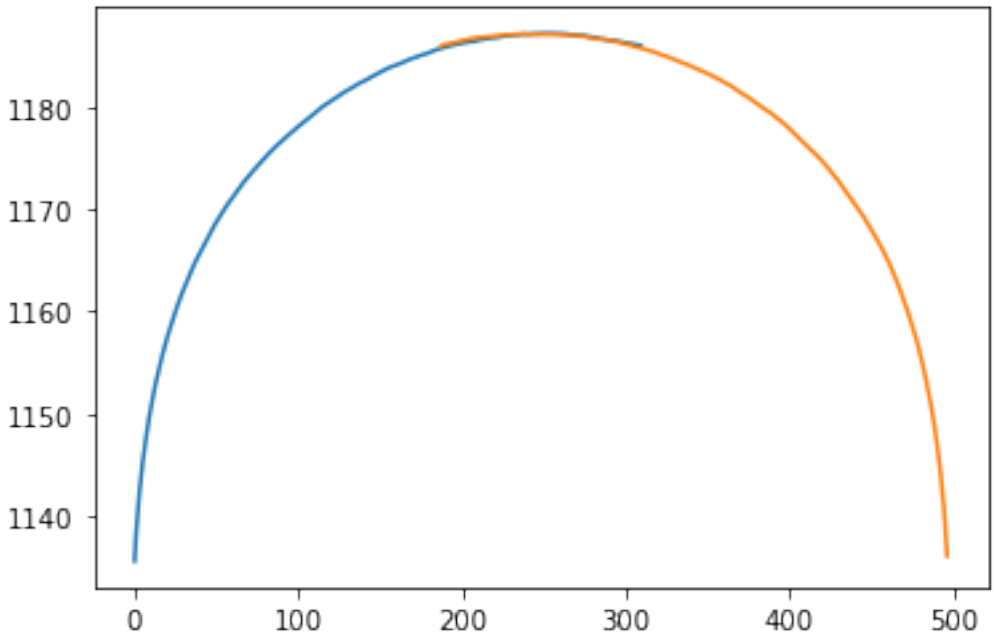
```
[(21351745, False), (21411078, False)]
```

26.2.2 Results

```
import matplotlib.pyplot as plt
```

```
N, M = len(system_params['StatisticalImage']['I0']), system_params['StatisticalImage']['M']
```

```
for i, r in enumerate(wlresults['results']):
    plt.plot(r['Es'][:-1], r['S'] + ΔS[i])
```



Fit a spline to interpolate and optionally clean up noise, giving WL g's up to a normalization constant.

```
gspl = interpolate.splrep(wlEs, S, s=0*np.sqrt(2))
wlgs = np.exp(interpolate.splev(wlEs, gspl) - min(S))
```

26.2.3 Exact solution

We only compute to halfway since g is symmetric and the other half's large numbers cause numerical instability.

```
def reflect(a, center=True):
    if center:
        return np.hstack([a[:-1], a[-1], a[-2::-1]])
    else:
        return np.hstack([a, a[::-1]])
```

The exact density of states for uniform values. This covers the all gray and all black/white cases. Everything else (normal images) are somewhere between. The gray is a slight approximation: the ground level is not degenerate, but we say it has degeneracy 2 like all the other sites. For the numbers of sites and values we are using, this is insignificant.

```
def bw_g(E, N, M, exact=True):
    return sum((-1)**k * special.comb(N, k, exact=exact) * special.comb(E + N - 1 - k*(M + 1), E
    ↪  - k*(M + 1), exact=exact)
        for k in range(int(E / M) + 1))
def exact_bw_gs(N, M):
    Es = np.arange(N*M + 1)
    gs = np.vectorize(bw_g)(np.arange(1 + N*M // 2), N, M, exact=False)
    return Es, reflect(gs, len(Es) % 2 == 1)
```

```
def gray_g(E, N, M, exact=True):
    return 2 * bw_g(E, N, M, exact=exact)
def exact_gray_gs(N, M):
    Es = np.arange(N*M + 1)
    gs = np.vectorize(gray_g)(np.arange(1 + N*M // 2), N, M, exact=False)
    return Es, reflect(gs, len(Es) % 2 == 1)
```

Expected results for black/white and gray.

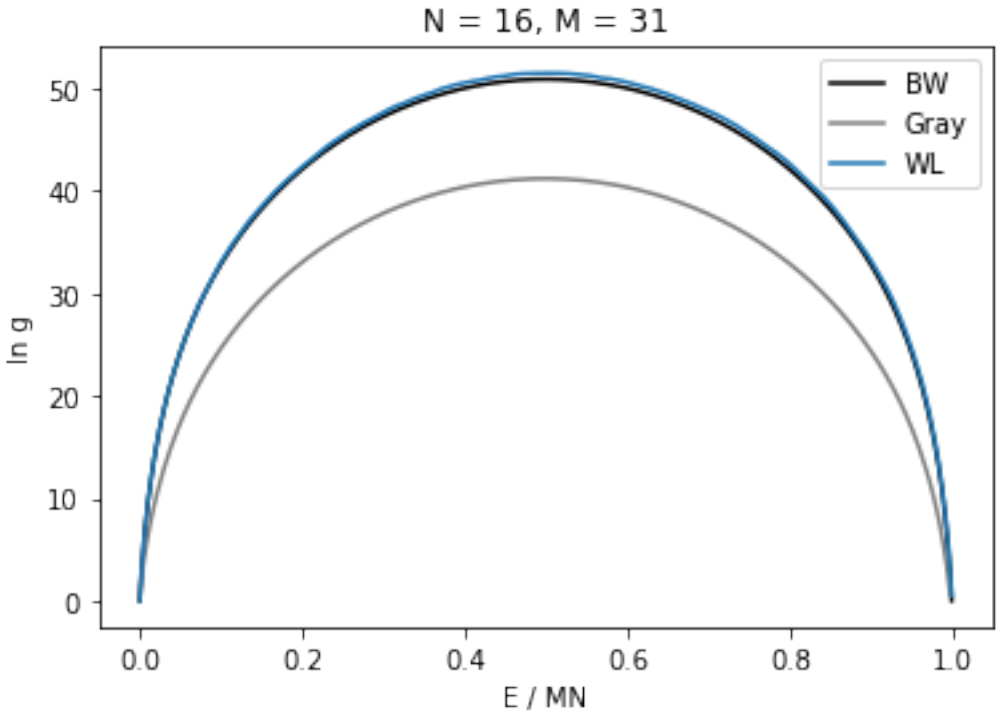
```
bw_Es, bw_gs = exact_bw_gs(N=N, M=M)
gray_Es, gray_gs = exact_gray_gs(N=N, M=-1 + (M + 1) // 2)
```

Choose what to compare to.

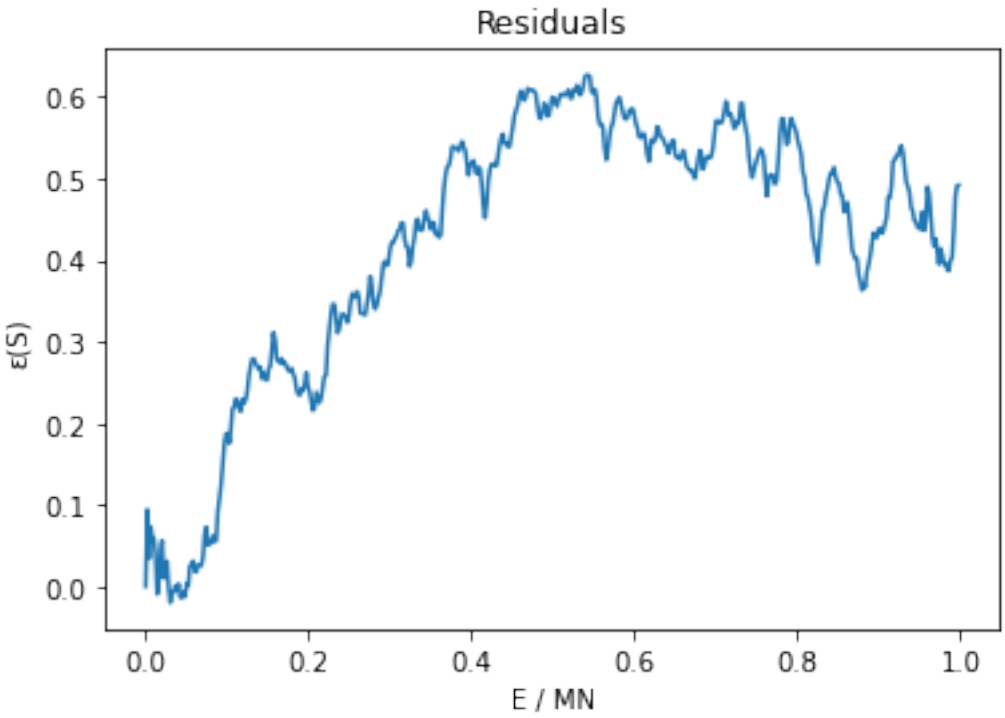
```
Es, gs = bw_Es, bw_gs
```

Presumably all of the densities of states for different images fall in the region between the all-gray and all-black/white curves.

```
plt.plot(bw_Es / len(bw_Es), np.log(bw_gs), 'black', label='BW')
plt.plot(gray_Es / len(gray_Es), np.log(gray_gs), 'gray', label='Gray')
plt.plot(wlEs / len(wlEs), np.log(wlgs), label='WL')
plt.xlabel('E / MN')
plt.ylabel('ln g')
plt.title('N = {}, M = {}'.format(N, M))
plt.legend();
```




```
# plt.plot(wlEs / len(wlEs), np.abs(wlgs - bw_gs) / bw_gs)
plt.title('Relative error')
plt.plot(wlEs / len(wlEs), S - np.log(bw_gs) - min(S))
plt.title('Residuals')
plt.xlabel('E / MN')
plt.ylabel('ε(S)');
```




```
print('End of job.')
```

```
End of job.
```

```
import numpy as np
import matplotlib.pyplot as plt
import os
import h5py, hickle
```

```
import sys
if 'src' not in sys.path: sys.path.append('src')
import wanglandau as wl
```

26.2.4 Calculating canonical ensemble averages

```
class CanonicalEnsemble:
    def __init__(self, Es, gs, name):
        self.Es = Es
        self.gs = gs
        self.name = name
    def Z(self, β):
```

```
        return np.sum(self.gs * np.exp(-β * self.Es))
    def average(self, f, β):
        return np.sum(f(self) * self.gs * np.exp(-β * self.Es)) / self.Z(β)
    def energy(self, β):
        return self.average(lambda ens: ens.Es, β)
    def energy2(self, β):
        return self.average(lambda ens: ens.Es**2, β)
    def heat_capacity(self, β):
        return self.energy2(β) - self.energy(β)**2
    def free_energy(self, β):
        return -np.log(self.Z(β)) / β
    def entropy(self, β):
        return β * self.energy(β) + np.log(self.Z(β))
```

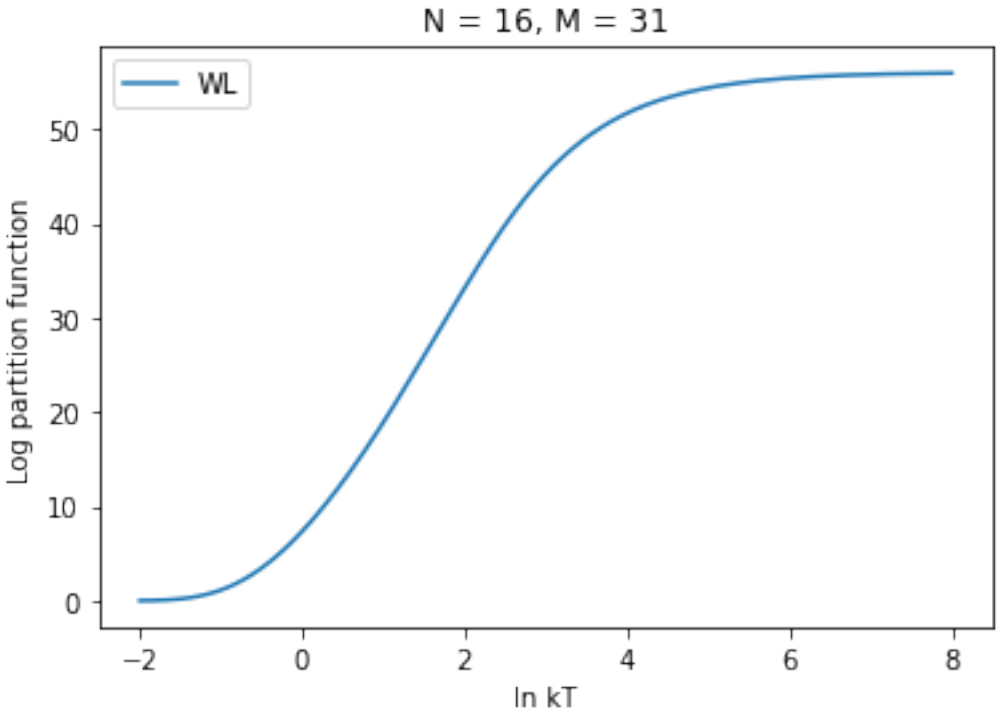
```
with h5py.File('data/simulation-l3t1t_sn.h5', 'r') as f:
    h = hickle.load(f)
    results = h['results']
    params = h['parameters']
```

```
N, M = len(params['system']['StatisticalImage']['I0']),
↪  params['system']['StatisticalImage']['M']
wlEs, S, ΔS = wl.join_results(results)
wlgs = np.exp(S - min(S))
```

```
βs = [np.exp(k) for k in np.linspace(-8, 2, 500)]
wlens = CanonicalEnsemble(wlEs, wlgs, 'WL') # Wang-Landau results
# xens = CanonicalEnsemble(Es, gs, 'Exact') # Exact
# ensembles = [wlens, xens]
ensembles = [wlens]
```

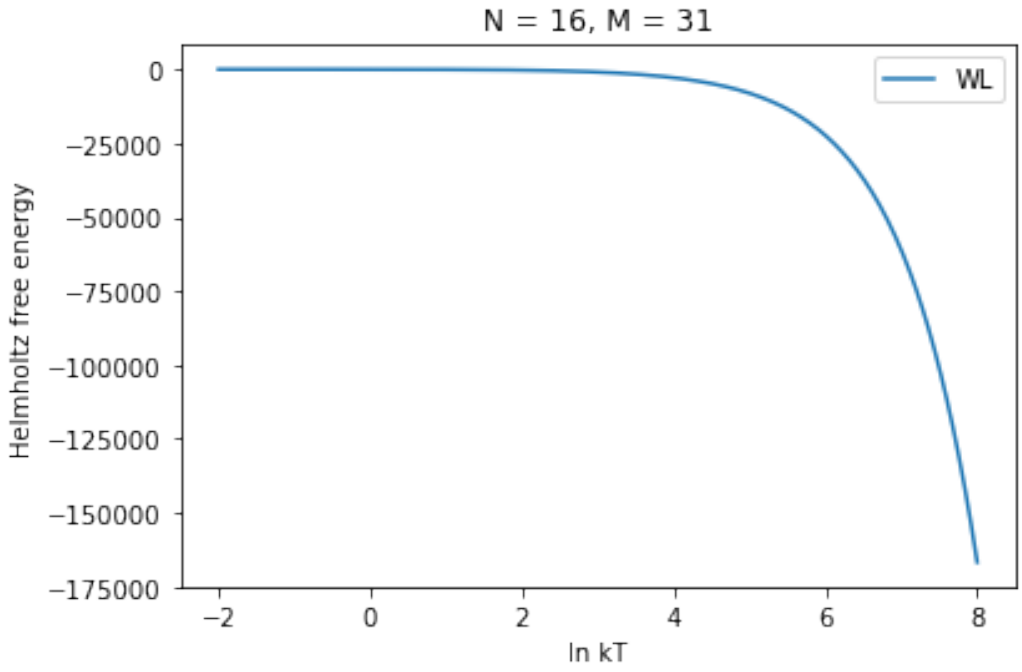
Partition function

```
for ens in ensembles:
    plt.plot(-np.log(βs), np.log(np.vectorize(ens.Z)(βs)), label=ens.name)
plt.xlabel("ln kT")
plt.ylabel("Log partition function")
plt.title('N = {}, M = {}'.format(N, M))
plt.legend();
```



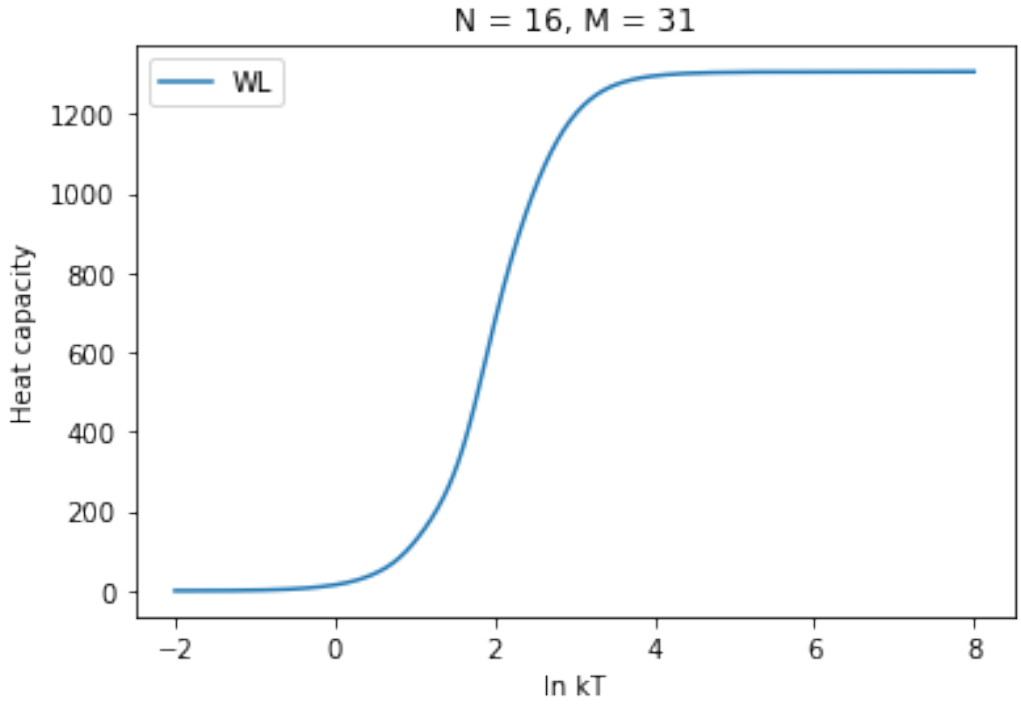
Helmholtz free energy

```
for ens in ensembles:
    plt.plot(-np.log(βs), np.vectorize(ens.free_energy)(βs), label=ens.name)
plt.xlabel("ln kT")
plt.ylabel("Helmholtz free energy")
plt.title('N = {}, M = {}'.format(N, M))
plt.legend();
```



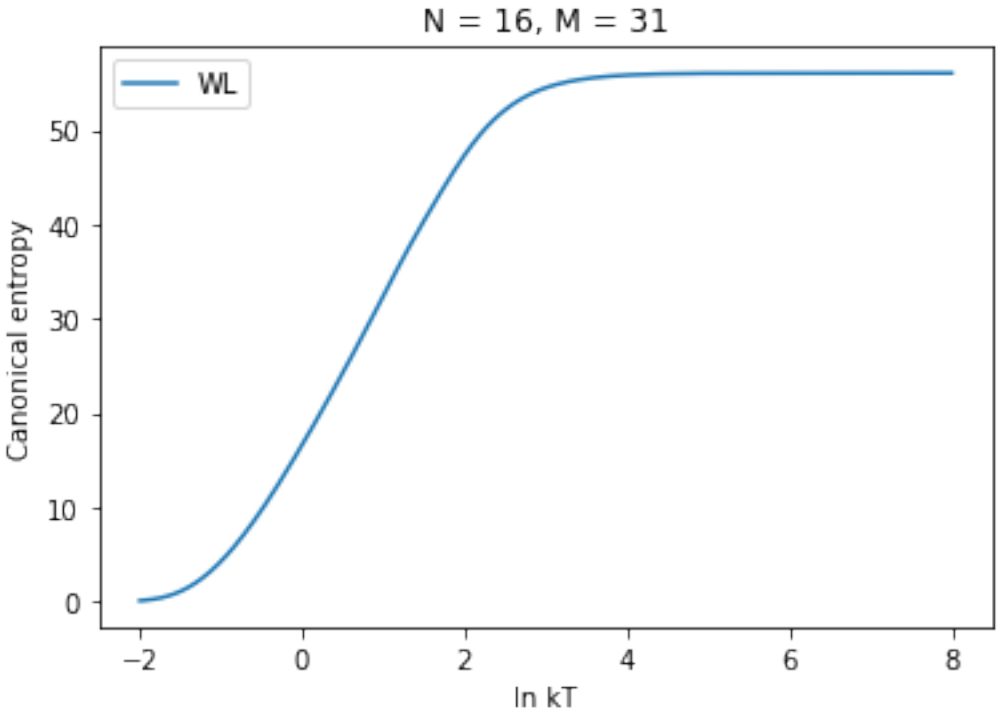
Heat capacity

```
for ens in ensembles:
    plt.plot(-np.log(βs), np.vectorize(ens.heat_capacity)(βs), label=ens.name)
plt.xlabel("ln kT")
plt.ylabel("Heat capacity")
plt.title('N = {}, M = {}'.format(N, M))
plt.legend();
```



Entropy

```
for ens in ensembles:
    plt.plot(-np.log(βs), np.vectorize(ens.entropy)(βs), label=ens.name)
plt.xlabel("ln kT")
plt.ylabel("Canonical entropy")
plt.title('N = {}, M = {}'.format(N, M))
plt.legend();
```



27 Comparison to Wang and Landau's results

June 18, 2020

28 Progress summary

June 19, 2020

This week, I further prepared to study the thermodynamics of our image systems. I derived the exact density of states for the all-black image, and the same technique (generating functions) could be used with computer enumeration of integer partitions to obtain the density of states for an arbitrary image system. I also spent more time improving my implementation of the Wang-Landau algorithm. It is now able to simulate arbitrary energy bins and divide different energy intervals across multiple CPU cores and combine them back together. The python code was made type-stable so that functions and the simulation state can be JITed by a LLVM-based compiler (Numba). These two improvements increased the speed of the simulations by more than a factor of 10. With some more tweaks and code to automate the preparation of parallel kernels and manage results, we should be able to easily perform these simulations for many parameter values

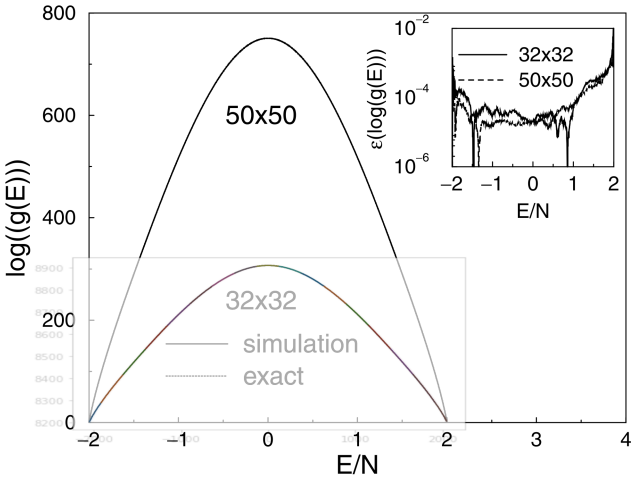


Figure 1: Our results agree with Wang and Landau's results, though it took more iterations than they reported to reproduce.

with high statistics.