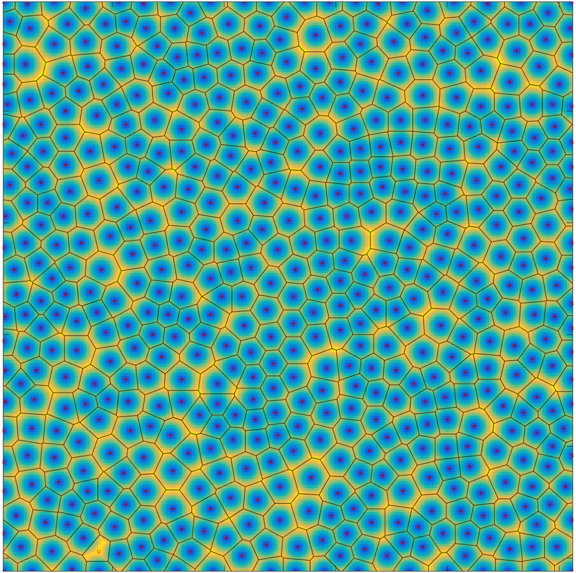
Farthest Point Sampling



Let $X^n = \{x_i\}_{i=1,\ldots,n}$ be a set of n points in some metric space, for example $x_i \in \mathbb{R}^N$, such that $d(x_i, x_j)$ can be computed. We would like to represent the large set X^n by a subset $Y^k = \{y_j\}_{j=1,\ldots,k}$, where $Y^k \subset X^n$ and $k \ll n$.
Denote the maximal distortion by

$$r^k = \max_{x_i \in X^n} \min_{y_j \in Y^k} d(x_i, y_j).$$

The optimal Y^k should satisfy

$$Y^{k*} = \arg_{Y^k \subset X^n} \min r^k.$$

That is, the k-quantization subset of X^n that minimizes the L_∞ norm. By definition, Y^{k*} yields the smallest possible r^{k*}.
Finding Y^{k*} is, generally speaking, NP hard.
Can we find a $Y^k \subset X^n$ for which the quantization error would be smaller than $\mathrm{c}r^{k*}$ for some positive constant c?
Such an approximation is known as c-approximation.

Farthest Point Sampling

Init: Let $y_1 = x_i$ for an arbitray $x_i \in X_n$, and let $Y = \{y_1\}$.

loop: Find the point $x_f \in X^n$ which is the farthest away from the selected points in Y. That is,

$$x_f = \max_{x_f \in X^n} \min_{y_j \in Y} d(x_f, y_j).$$

- Let $Y = Y \cup \{x_f\}$
- If $|Y| < k$ go to loop.

Denote the output as $Y^{k_{FPS}}$.

Next, we will prove that this simple greedy strategy is a 2-approximation.

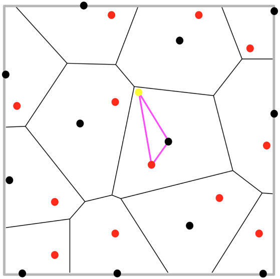
Proof of 2-Approximation

Proof sketch.
Let v_j be the Voronoi cell about $y_j^* \in Y^{k*}$.
v_j includes all points in X_n whose distance to y_j^* is the smallest.
That is, $x_i \in v_j$ if $d(x_i, y_j^*) \leq d(x_i, y_l^*)$ for all $y_l^* \in Y^{k*}$, such that $l \neq j$.
Using the pigeon hole principle, the proof shows that either each point in $Y^{k_{FPS}}$ falls into a different Voronoi cell of Y^{k*}, or there is at least one case in which there are two (or more) points of $Y^{k_{FPS}}$ that fall into the same Voronoi cell of Y^{k*}. In both cases $r^k \leq 2r^{k*}$.

Case 1: Assume that each $y_j \in Y^{k_{FPS}}$ falls into a different Voronoi cell. That is, at each v_j with representative point $y_j^* \in Y^{k*}$, there is exactly one y_j. By definition, $d(y_j, y_j^*) \leq r^{k*}$. Next, for each $x_i \in v_j$ we have $d(x_i, y_j^*) \leq r^{k*}$. Since we deal with a metric space, by the triangle inequality, for each $x_i \in v_j$ we have,

$$d(x_i, y_j) \leq d(x_i, y_j^*) + d(y_j^*, y_i) \leq 2r^{k*}.$$

Repeating for all points in X^n, each point within its own cell, concludes this part.

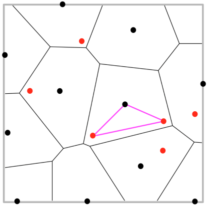


Case 2: Assume that at step $l \leq k$ of the FPS procedure, two points y_m and y_l, where $m < l$, fall into the same Voronoi cell v_j about $y_j^* \in Y^{k*}$. The maximal distortion of the FPS strategy is $r^l \leq r^m$, for all $m < l$. In fact, the closest point in Y^{l-1} to the newly selected y_l is at distance

$$r^{l-1} = \max_{x_i \in X^n} \min_{y_j \in Y^{l-1}} d(x_i, y_j),$$

which is how the FPS selects its candidates from $X^n \setminus Y^{l-1}$. We need to show that if y_l and y_m both fall into v_j, then $r^{l-1} \leq 2r^{k*}$. By construction, $r^{l-1} \leq d(y_l, y_m)$. We thus have that

$$r^k \leq r^{l-1} \leq d(y_l, y_m) \leq d(y_l, y_j^*) + d(y_j^*, y_m) \leq 2r^{k*}$$



□

Matlab Example

```
N = 256; A = zeros(N);
X = 1+floor(rand(1)*N);
Y = 1+floor(rand(1)*N);
A(X,Y) = 1;
for k = 1:1000,
    figure(1);
    B = bwdist(A);
    imagesc(B'); axis image;hold on;
    mx = max(B(:)); ind = (B == mx);
    [I,J] = find(ind,1);
    A(I,J)=1;
    X = [X; I]; Y = [Y; J];
    if max(size(X))>4,
        [vx,vy] = voronoi(X,Y);
        plot(X,Y,'ro',vx,vy,'k-');
        axis equal;axis([1 N 1 N]);
        hold off;
        drawnow;
    end %if
end
```