Adaptive thinning of centers for approximation by radial functions

Nira Dyn
School of Mathematical Sciences
Tel Aviv University, Israel

Joint work with Pavel Kozlov (M.Sc thesis)

September 2013

Outline of the Talk

1. The approximation problem

2. Adaptive thinning algorithms

3. An anticipated error functional

4. Some predicting functionals

5. Heuristic explaination

6. Numerical examples

The approximation problem

The data

- a finite set of distinct centers (points) $\Xi \subset \mathbb{R}^d$
- function's values at these centers $\{F(\xi)) : \xi \in \Xi\}$.
- a prescribed error bound ϵ

The problem: For a given radial function $\varphi : \mathbb{R}_+ \to \mathbb{R}$, to find a small subset of Ξ, Y, such that the best ℓ_2-approximation to F on Ξ from span$\{\varphi(\| \cdot - y\|) : y \in Y\}$, $S(Y, \varphi)$, satisfies

$$\|(F - S(Y,\varphi)\|_{\ell_2(\Xi)} = \left(\frac{1}{|\Xi|}\sum_{x\in\Xi}(F(x) - S(Y,\varphi)(x))^2\right)^{\frac{1}{2}} \leq \epsilon$$

In this talk we show that the problem we stated is feasibile, and we give a method for its solution. We do not have estimates relating the number of centers needed for a certain accuracy with the properties of the approximated function.

Such theoretical results are obtained in a paper by Devore and Ron (2008), where a method for placement of centers is studied. The method is based on the expansion of the approximated function by wavelets, and on the approximation of the wavelets by translates of a radial function

The method of solution–Adaptive thinning

- Removal of least significant centers, one by one in a *greedy way* ([Dyn, Floater, Iske (2000)]).

- For a set of centers Y and an anticipated error functional $e(y;Y,\varphi)$, estimating the error incurred by the removal of y from Y, the center with least anticipated error is the least significant.

- The novelty in our approach is the use of a **predicting functional** instead of an anticipated error functional.

- The functional $p(y;Y,\varphi)$ is a predicting functional for $e(y;Y,\varphi)$ if it determines with high probability the same least significant center as $e(y;Y,\varphi)$.

  We call $p(y;Y,\varphi)$ the **significance** of y in Y relative to p.

Adaptive thinning algorithm

- Set $Y = \Xi$

- Compute $S(Y, \varphi)$

- While $\|(F - S(Y,\varphi))\|_{\ell_2(\Xi)} \leq \epsilon$

  1. compute the significance of each y in Y.
  2. find y^*–the least significant center in Y.
  3. set $Y = Y \setminus y^*$.
  4. compute $S(Y, \varphi)$

- Set $Y = Y \cup y^*$, and return Y as the set of significant centers

From the true error to an anticipated error

For $Y \subseteq \Xi$, let $S(Y,\varphi) = \sum_{y\in Y} \alpha_y \varphi(\|\cdot - y\|)$

A heuristic argument

The error incurred by the removal of $y \in Y$ from Y, $E(y;Y,\varphi) = \|F - S(Y \setminus y, \varphi)\|_{\ell_2(\Xi)}$, satisfies

$$\|F - S(Y,\varphi)\|_{\ell_2(\Xi)} \leq E(y;Y,\varphi) \leq \|F - (S(Y,\varphi) - \alpha_y \varphi(\|\cdot - y\|))\|_{\ell_2(\Xi)}$$

For a center of small significance y we can assumne that the upper and lower bounds above are close

Thus in case $\{\alpha_y : y \in Y\}$ are known, an anticipated error functional is

$$e(y;Y,\varphi) = \|F - \sum_{z\in Y\setminus y} \alpha_z \varphi(\|\cdot - z\|)\|_{\ell_2(\Xi)}$$

From the anticipated error to a predicting functional

From $e(y; Y, \varphi)$ we can derive a predicting functional, in view of the following proposition

Proposition

$$\|F - \sum_{z \in Y \setminus y} \alpha_z \varphi(\| \cdot - z\|)\|^2_{\ell_2(\Xi)} = \|F - S(Y, \varphi)\|^2_{\ell_2(\Xi)} + \|\alpha_y \varphi(\| \cdot - y\|)\|^2_{\ell_2(\Xi)}$$

The functional $p(y; Y, \varphi) = |\alpha_y| \|\varphi(\| \cdot - y\|)\|_{\ell_2(\Xi)}$ is a predicting functional, since

$$\arg\min_{y \in Y} p(y; Y, \varphi) = \arg\min_{y \in Y} e(y; Y, \varphi)$$

but $p(y; Y, \varphi)$ is not an estimate of the error incurred by the removal of y from Y.

Simplifying the predicting functional

Although the computation of $p(y;Y,\varphi) = |\alpha_y| \|\varphi(\| \cdot - y\|)\|_{\ell_2(\Xi)}$ has a lower complexity than the computation of the true error $E(y;Y,\varphi) = \|F - S(Y \setminus y, \varphi)\|_{\ell^2(\Xi)}$, this complexity is still high

Next we simplify $p(y;Y,\varphi)$ for positive, strictly monotone radial functions

For given $\{\alpha_y \ : \ y \in Y\}$ we search for a simpler to compute functional $\lambda(y;Y,\varphi)$ for which the equality

$$\arg\min_{y \in Y} p(y;Y,\varphi) = \arg\min_{y \in Y} \lambda(y;Y,\varphi)$$

holds with *high probability*

Two observations allow us to obtain simpler predicting functionals.

First obsevation–consistency of functionals

Let $B(\mathbf{0}, R)$ denote the ball with center at the origin and radius R, let $y, z \in B(\mathbf{0}, R)$, and let φ be a positive radial function which is strictly monotone on $[0, 2R]$. Then the following three statements are equivalent:

(i) $\|y - \mathbf{0}\| > \|z - \mathbf{0}\|$

(ii) Let $\sigma = +1(-1)$ for φ increasing (decreasing). Then for $p \in [1, \infty)$

$$\sigma \|\varphi(\| \cdot - y\|)\|_{L_p(B(\mathbf{0},R))} > \sigma \|\varphi(\| \cdot - z\|)\|_{L_p(B(\mathbf{0},R))},$$

(iii) With σ as above and

$$\mu(f) = \begin{cases} \max_{x \in B(\mathbf{0},R)} f(x) & \text{for } \varphi \text{ increasing} \\ \min_{x \in B(\mathbf{0},R)} f(x) & \text{for } \varphi \text{ decreasing} \end{cases}$$

$$\sigma \mu(\varphi(\| \cdot - y\|)) > \sigma \mu(\varphi(\| \cdot - z\|))$$

A heuristic conclusion

Let φ be a positive, strictly monotone radial function, and let the set Ξ of centers be "nicely distributed" in a "nice" domain. We assume that with *high probability*

$$\|\varphi(\|\cdot - y\|)\|_{\ell_2(\Xi)} > \|\varphi(\|\cdot - z\|)\|_{\ell_2(\Xi)},$$

if and only if:

for φ increasing

$$\max_{x\in\Xi} \varphi(\|x-y\|) > \max_{x\in\Xi} \varphi(\|x-z\|)$$

and for φ decreasing

$$\min_{x\in\Xi} |\varphi(\|x-y\|) > \min_{x\in\Xi} \varphi(\|x-z\|)$$

Note that a similar equivalence also holds with the above three inequality signs replaced by three equality signs.

The Heuristic conclusion leads us to replace the predicting functional

$$p(y; Y, \varphi) = |\alpha_y| \|\varphi(\| \cdot - y\|)\|_{\ell_2(\Xi)}$$

by the simpler functional

$$\lambda(y; Y, \varphi) = |\alpha_y| \mu(\varphi(\| \cdot - y\|))$$

with $\mu(f) = \bar{\mu}(f) = \max_{x \in \Xi} f(x)$ for φ increasing, and
with $\mu(f) = \underline{\mu}(f) = \min_{x \in \Xi} f(x)$ for φ decreasing.

Inconsistency happens when either

$$p(y; Y, \varphi) > p(z; Y, \varphi) \quad \text{and} \quad \lambda(y; Y, \varphi) < \lambda(z; Y, \varphi)$$

or when

$$p(y; Y, \varphi) < p(z; Y, \varphi) \quad \text{and} \quad \lambda(y; Y, \varphi) > \lambda(z; Y, \varphi)$$

In the first case, the ratio $\frac{\alpha_y}{\alpha_z}$ is confined to the interval

$$I(y, z) = (a(y, z), b(y, z)) = \left(\frac{\|\varphi(\| \cdot - z\|)\|_{\ell_2(\Xi)}}{\|\varphi(\| \cdot - y\|)\|_{\ell_2(\Xi)}}, \frac{\mu(\varphi(\| \cdot - z\|)}{\mu(\varphi(\| \cdot - y\|)} \right)$$

In the second case $a(y,z) > b(y,z)$ and the ratio $\frac{\alpha_y}{\alpha_z}$ is confined to the interval $(b(y,z), a(y,z))$, which we also denote by $I(y,z)$

We call the interval $I(y,z)$ *inconsistency interval*

It is sufficient to consider all pairs of distinct points of Y in the set $Y^2_> = \{(y,z) \in Y \times Y : \ \|\varphi(\|\cdot - y\|)\|_{\ell_2(\Xi)} > \|\varphi(\|\cdot - z\|)\|_{\ell_2(\Xi)}\}$

Note that $I(y,z) \subset (0,1)$ for $(y,z) \in Y^2_>$, if there is functional consistency between μ and $\|\cdot\|_{\ell_2(\Xi)}$

Our second observation estimates the probability of the ratio $\frac{\alpha_y}{\alpha_z}$ to be in an inconsistency interval, under reasonable assumptions. We checked numerically that this probability is small for the two radial functions we work with

$$\varphi(r) = r^3 \quad \text{and} \quad \varphi(r) = \exp(-0.1r)$$

Second observation- inconsistency due to $\{\alpha_x : \; x \in \Xi\}$

Reasonable Assumptions (for a large set Ξ, under the lack of information about the distribution of the ratios $\{|\frac{\alpha_y}{\alpha_z}| : \; (y,z) \in \Xi^2_>\}$ in (0,1))

(i) The inconsistency intervals are contained in $(0,1)$

and therefore if $|\alpha_y| \geq |\alpha_z|$ there is no inconsistency

(ii) For $|\alpha_y| < |\alpha_z|$ the ratio $|\frac{\alpha_y}{\alpha_z}|$ is uniformly distributed in the interval $(0,1)$

(iii) For any set of coefficients $\{\alpha_x : \; x \in \Xi\}$, and for any $(y,z) \in \Xi^2_>$, the probability that $|\frac{\alpha_y}{\alpha_z}| < 1$ equals the probability that $|\frac{\alpha_y}{\alpha_z}| > 1$

It follows from the above assumptions that the probability of a ratio $|\frac{\alpha_y}{\alpha_z}|$ for $(y,z) \in \Xi^2_>$ to be contained in an inconsistency interval equals half times the length of $I(y,z)$

The length of an inconsistency interval is

$$L(y,z) = |b(y,z) - a(y,z)| = \left| \frac{\|\varphi(\|\cdot - z\|)\|_{\ell_2(\Xi)}}{\|\varphi(\|\cdot - y\|)\|_{\ell_2(\Xi)}} - \frac{\mu(\varphi(\|\cdot - z\|)}{\mu(\varphi(\|\cdot - y\|)} \right|$$

The probability of inconsistency due to the coefficients $\{\alpha_x : \ x \in \Xi\}$ is half times the average length of the inconsistency intervals corresponding to pairs of points in

$$\Xi^2_> = \{(y,z) \in \Xi \times \Xi : \ \|\varphi(\|\cdot - y\|)\|_{\ell_2(\Xi)} > \|\varphi(\|\cdot - z\|)\|_{\ell_2(\Xi)}\}$$

$$P_{inc}(p, \lambda; \Xi) = \frac{1}{2|\Xi^2_>|} \sum_{(y,z) \in \Xi^2_>} L(y,z)$$

Our aim is to obtain a computable apreiori estimate of the probabilty of inconsistency caused by the coefficients $\{\alpha_x : \ x \in \Xi\}$

For a large set of points Ξ, which are "nicely" distributed in a domain D, we estimate the average length of the inconsistensy intervals by replacing the sum appearing in $P_{inc}(p, \lambda; \Xi)$ by an integral

$$P_{inc}(p, \lambda; D) = \frac{1}{2|DD_>|} \int\limits_{DD_>} L(y, z) \mathrm{d}y \mathrm{d}z$$

with

$$DD_> = \{(y, z) \in D \times D : \; \|\varphi(\| \cdot - y\|)\|_{L_2(D)} > \|\varphi(\| \cdot - z\|)\|_{L_2(D)}\}$$

The quality of $P_{inc}(p, \lambda; D)$ as an estimate of the probability of inconsistency for subsets Y of Ξ deteriorates as the size of Y decreases

Numerical observations

$$D = [-1, 1],\ \varphi(r) = r^3,\ P_{inc}(p, \bar{\mu}; D) \approx 0.009$$

$$D = [-1, 1],\ \varphi(r) = \exp(-0.1r),\ P_{inc}(p, \underline{\mu}; D) \approx 0.004$$

For $\varphi(r) = \exp(-0.1r)$

$$\max_{y \in B(\mathbf{0},R)} \underline{\mu}\varphi(\|\cdot - y\|) - \min_{y \in B(\mathbf{0},R)} \underline{\mu}\varphi(\|\cdot - y\| = \varphi(R) - \varphi(2R)$$

$$\max_{R>0}(\varphi(R) - \varphi(2R)) = 0.25 \text{ attained at } R = 6.9$$

$$\varphi(1) - \varphi(2) = 0.086 \text{ and } \varphi(100) - \varphi(200) = 0.000045$$

implying that $\underline{\mu}(\|\cdot - y\|)$ is almost a constant for $y \in D$, where D is a "nice" large domain

The last numerical observations suggest that the functional $\underline{\mu}$ can be replaced by the functional $\mathbf{1}(f) = 1$ in the predicting functional λ. Thus $|\alpha_y|$ is a predicting functional for this φ

The "best" predicting functional

Our numerical tests indicate that the predicting functional

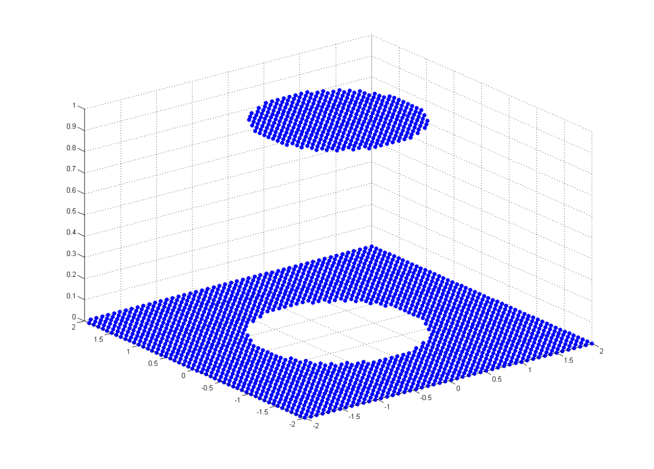
$$p^*(y;Y,\varphi) = |\alpha_y \varphi(\min_{z\in Y\setminus y} \|z-y\|)|$$

works very well for any radial function

Our efforts to explain this "magic" lead us to the predicting functionals we discussed before

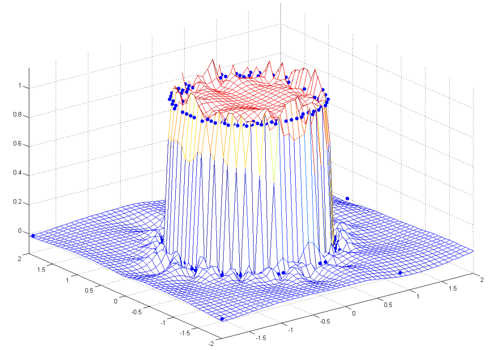
ATR2 and ATRM($\bar{\mu}$) $\varphi(r) = r^3, \epsilon = 0.1$

2500 samples of a cylinder function

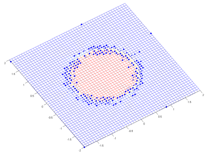


ATR2 and ATRM($\bar{\mu}$) $\varphi(r) = r^3, \epsilon = 0.1$

ATR2: 287 centers, 2500 data



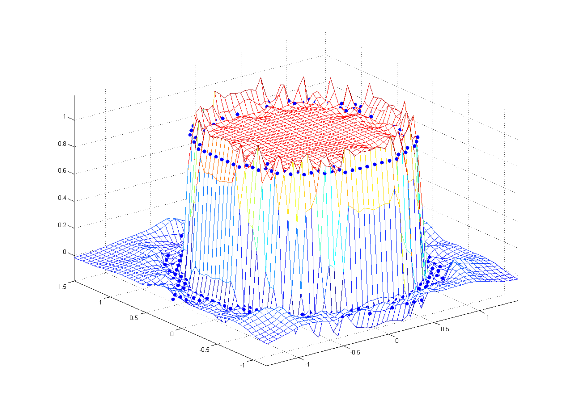
significant centers.



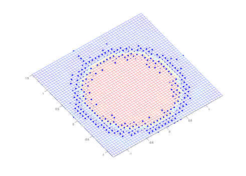
Note that most of the significant centers are near discontinuity points.

ATR2 and ATRM($\bar{\mu}$) $\varphi(r) = r^3, \epsilon = 0.1$

ATRM($\bar{\mu}$): 381 centers, 2500 data

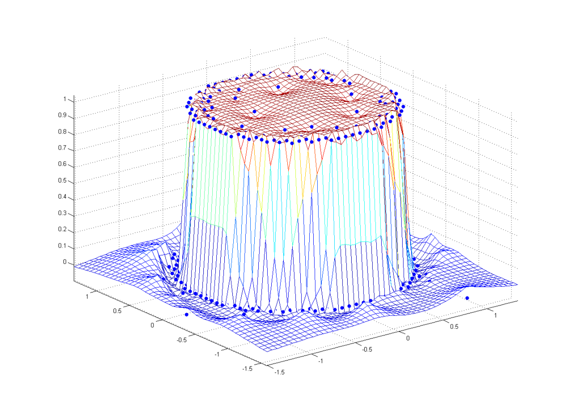


significant centers.

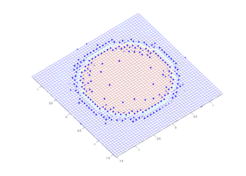


ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}, \epsilon = 0.1$

ATR2: 362 centers, 2500 data

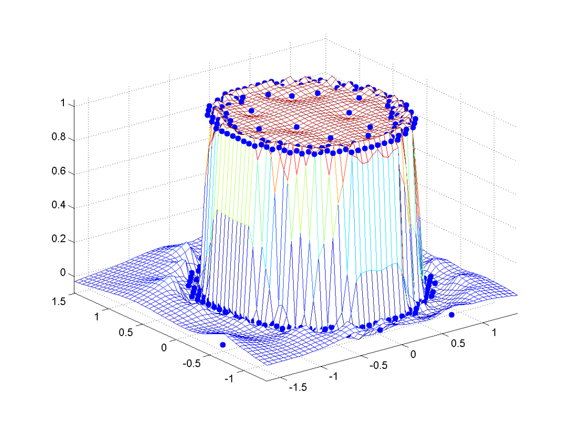


significant centers.

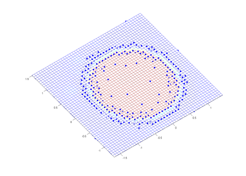


ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}, \epsilon = 0.1$

ATRM(1): 377 centers, 2500 data

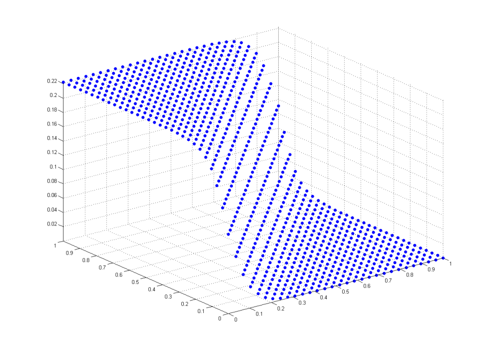


significant centers.



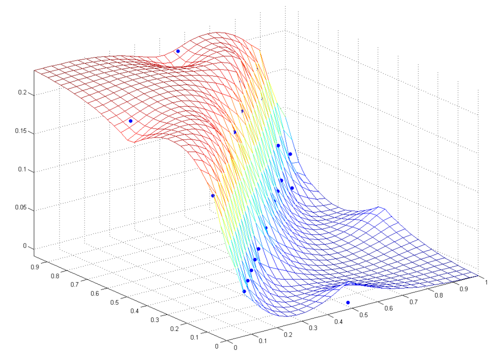
ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}, \epsilon = 0.01$

900 samples of a test function from Dyn,Levin,Rippa (1990)

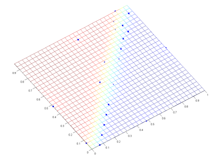


ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}$, $\epsilon = 0.01$

ATR2: 29 centers, 900 data



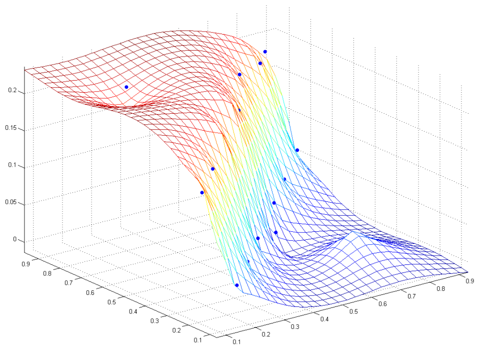
significant centers.



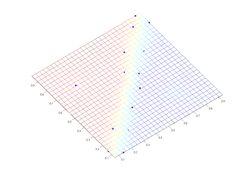
Note that many significant centers are located along the line of large gradient.

ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}, \epsilon = 0.01$

ATRM(1): 17 centers, 900 data

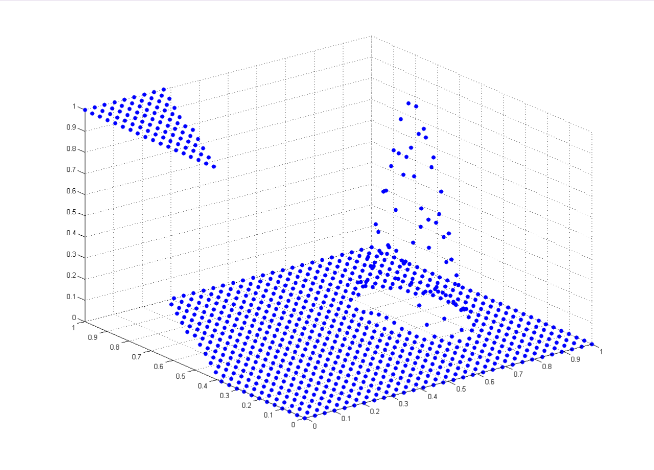


significant centers.



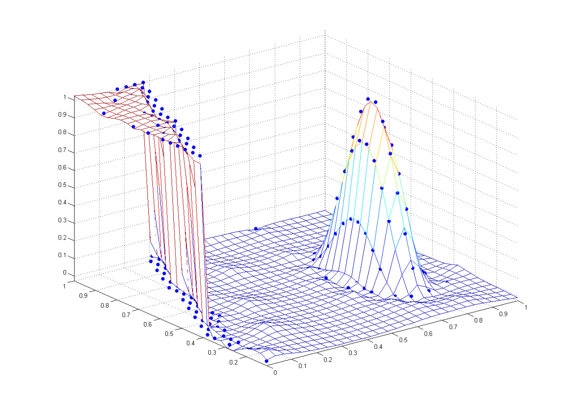
ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}, \epsilon = 0.01$

900 samples of Ritchie's (1978) function

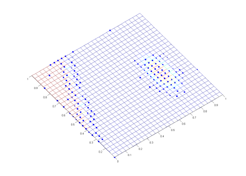


ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}, \epsilon = 0.01$

ATR2: 195 centers, 900 data



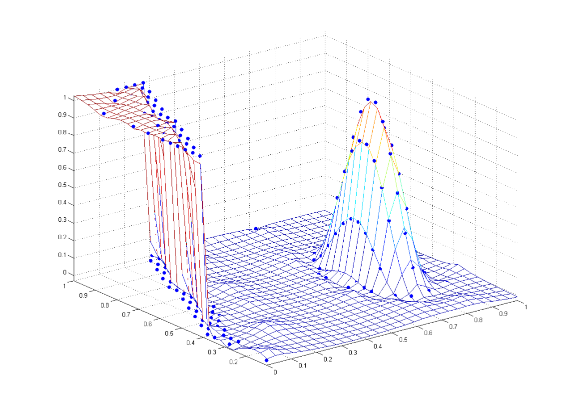
significant centers.



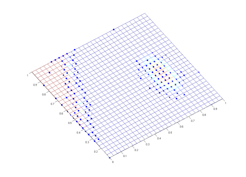
Note that most of the significant centers are located near points of discontinuity or points of large gradient.

ATR2 and ATRM(1) $\varphi(r) = e^{-0.1r}$, $\epsilon = 0.01$

ATRM(1): 197 centers, 900 data



significant centers.



Summary

Our experiments indicate that for a fixed level of error:

- ATR0 selects the smallest number of significant centers but it has the highest computational cost.
- ATR2 is close to ATR0. ATR2 outperforms ATRM($\bar{\mu}$) for $\varphi \uparrow$. ATRM($\mathbf{1}$) is close to ATR0 and it is close or outperforms ATR2 for $\varphi \downarrow$ and has the lowest computational cost.
- The ATRM algorithms are based on our good heuristic and on the Functional Consistency Theorem.
- We have no explanation for the success of ATR2.