On some insight and extensions of the Radial Basis Function method.

Andrea Cassioli$^1$

$^1$LIX - ECOLE POLYTECHNIQUE (FR)
cassioli@lix.polytechnique.fr

11/07/2013
Outline

1. Brief summary of the RBF method
2. Some more insight on the bumpiness
3. Other variants of the RBF method
The RBF method

Problem definition

We aim to solve

$$\min_{x \in X} f(x)$$

where:

- $X \subset \mathbb{R}^n$ is the (bounded) feasible set of $x$
- $f(x)$ is a black-box function whose evaluation is “costly”
For a given (unisolvent) set of samples $S$:

$$s(x|S) = \sum_{y \in S} \lambda_y \phi(\|x - y\|) + p(x, c) = \Phi(x)\lambda + P(x)^Tc$$

Coefficients $\lambda, c$ are computed solving the linear system

$$\Phi\lambda + Pc = f$$

$$P^T\lambda = 0$$

the degree of the polynomial depends on $\phi()$. 

Cassioli (LIX)
The RBF method

Basic algorithm

determine suitable $S_0$;
$k \leftarrow 0$;

while stopping criteria not fulfilled do

  determine $s_k(x|S_k)$;

  $y_k = \arg \min_{x \in X} s_k(x)$;

  select the aspiration level $\hat{f}$;

  determine $x_{k+1}$ based on $\mu(x|y_k, \hat{f})$;

  $S_{k+1} = S_k \cup x_k + 1$;

end
The RBF method

Comments

A lot of freedom:

- which radial basis
- the degree of the polynomial
- how solve the auxiliary problems
- how to select the reference value
The RBF method
Do we really care about convergence?

"In principle, these methods may have convergence guarantees if the point selection strategy is well-chosen; but this is irrelevant in view of the fact that for expensive functions, only few (perhaps up to 1000) function evaluations are admissible"\(^a\)

The RBF method
Convergence

Based on the well known theorem of Törn, A., Zilinskas, A.:

**Theorem**

*If an algorithm generates a sequence of points that are dense in in the feasible set X it converges to the optimal solution.*

Basically we will get arbitrary close to optimum...
The RBF method
Convergence

Theorem
If an RBF method is well posed (see usual properties of $s()$) and

- $S_k$ is unisolvent
- for the reference value holds that
  \[ \hat{f}_{k+1} < \min_{x} s(x|S_k) \]
- $x_{k+1}$ is a minimizer (maximizer) for the bumpiness function

then the point selected at iteration $k + 1$ is distinct for any other points in $S_k$. 

Cassioli (LIX)
Corollary

*If for the designed RBF method the previous theorem holds, then the method converges to the global optimum of $f(x)$.*

Proof.

If the previous theorem holds, then the sequence of points $\{x_i\}$ is dense in $X$ and for the Torn and Zilinskas theorem we converge to the global optimum.
The RBF method
Convergence

- Granted if an infinite subsequence of sampled point is dense in $X$
- In some cases convergence to first-order stationary points
- In probability if we can sample the feasible set along $\{x_k\}$
Let consider 1D cubic splines for a set \( \{x_1, \ldots, x_k\} \):

- must fulfill some condition on first/second derivatives
- it's *natural* if \( s''(x) = 0 \)
- minimize

\[
I(s) = \int_{-\infty}^{\infty} s''(x)^2 \, dx
\]

which a curvature measure

The 1D cubic RBF is a natural spline

\[
s(x) = \sum_{i=1}^{k} \lambda_i |x - x_i|^3 + c_1 + c_2 x
\]
On the bumpiness function
A step back on the interpolant function

Generalize to a general radial basis $\phi()$, we obtain

$$I(s) = \int_{-\infty}^{\infty} s''(x)^2 \, dx = \ldots = 12\lambda^T \Phi \lambda + 12P\lambda,$$

but asking for a natural spline we get

$$I(s) = 12\lambda^T \Phi \lambda$$

An 1D RBF is the natural spline for that basis and set of points.
Moving to the multidimensional case, we note that the $I(s)$ function comes from the product

$$< u(x, \lambda), v(x, \mu) > = (-1)^m \sum_{i=1}^{k} \lambda_i v(x_i) = (-1)^m \sum_{i=1}^{k} \mu_i u(x_i),$$

yielding

$$< s, s >= (-1)^m \lambda^T \Phi \lambda$$

which is a seminorm once $P^T \lambda = 0$. 
On the bumpiness function

Meaning

For the surrogate model, *centers* are fixed, and we look for the λ’s.

For the *bumpiness*, one center is not fixed (the next point) and we minimize the seminorm of

\[ s(x|S \cup \hat{x}) = s(x|S) + (\hat{f} - s(\hat{x}|S))\hat{L}(x) \]

where \( L \) is an RBF of the same family that attains 1 in \( \hat{x} \) and 0 everywhere else.
On the bumpiness function

Meaning

How to think about the bumpiness?
Imagine:

▶ an elastic carpet that has be fixed in points at certain heights
▶ put your finger at the aspiration level
▶ move it until you find the point in which the carpet resists less to your pressure
▶ this is the next point!
On the bumpiness function
Pros...

▶ "simple" method
▶ a meaningful concept
▶ promote convergence
▶ allow for balancing exploration/intensification via the aspiration level
On the bumpiness function

...and Cons

- hard to optimize (very bumpy...)
- numerically unstable (log scaling)
- boundary "effect"
- difficult to relate to the geometry of $S$
- requires a (good) lower bound on the optimal value of the surrogate model
**CORS-RBF**

1. Use the surrogate model has merit function solving

\[
\min_{x \in X} s(x) \\
\|x - x_i\| \geq \Delta_i \\
i = 1 \ldots k
\]

2. Parallel version

---

next point among a pool of perturbations of the best solution so far
putative points scored using $s()$ and/or the geometry of the sample set
several variants depending on the scoring and globalization strategies
convergence in probability

---

ConstrLMSRBF$^4$

1. extends the LMSRBF version of SRS
2. build surrogates models for both objective function and constraints
3. require a first feasible point
4. consider feasibility violation in scoring the putative points
5. tested up to 4000 function evaluations

A Trust-Region based RBF method (no bumpiness):

1. test the model for "validity" and add new points if necessary
2. find a minimizer of the model in the TR
3. compute the improvement ratio
4. update TR

---

Extends BOOSTER with a more complex handling of the TR.

1. use only a subset of samples
2. the surrogate is built enforcing well conditioning (fully linearity)
3. the next point is the (approximate) minimizer of the surrogate on the TR
4. very complex framework

---


extends the original RBF method

consider BB constraints as penalty

select next point using bumpiness

aspiration level is varied and putative next points clustered

in some cases the aspiration level is ignored and the optimum of $s()$ is used

---

At iteration $k$, the Grid Mode uses a set $w = \{w_1 \ldots w_t\}$ of positive weight and determines

$$x_i = \arg \min_{x \in X} \mu(x, s_k - w_i f_{\Delta}) \quad i = 1 \ldots t,$$

points are then clustered$^9$ and one is selected using heuristics.

use an alternative merit function

consider approximation and interpolation

extends to multi-objective optimization

---

From interpolation to approximation $\eta[0, 1]$:

$$\min \quad \eta \lambda^T \Phi \lambda + (1 - \eta) \|\epsilon\|^2$$

s.t.

$$\Phi \lambda + P \epsilon = \epsilon + f$$

$$P^T \lambda = 0$$

$$\epsilon \in \mathbb{R}^k$$

1. $\eta \rightarrow 0$ yields original RBF method
2. $\eta \rightarrow 1$ yield the smoothest surrogate model

The choice of $\eta$ can be done using cross-validation.
It maximizes

\[ Q(y) = \int_{\Omega} (U_S(x) - U_{S\cup y}(x))\omega(s(x|S))dV(x) \]

where \( \omega() \) is a suitable weight function and

\[ U(x) = \min_{z \in S} \| x - z \| \]
References I


References III

