"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.6, 33-44 | Science Signpost Publishing

Vanishing Hermite Polynomials in Global Optimization

George A. Vazmin

Abstract

Weighted Hermite polynomials row for function symmetrized around objective (unknown) point opens a way for a localization of the global maximum position. This approach creates subsidiary function guiding to the most prominent extrema (up to 3). The guidance function is generalized to multi-dimensional space. Several analytical tools (suppression of noise, space constriction, etc) are described in this technology.

Keywords: Hermite polynomials, symmetrization, guidance, multi-dimensional, noise suppression, scaling, space collapse, recursive iterations, poor statistics.

Hermite Polynomials Series for Symmetrized Function

The rising importance in analysis of multi-extreme functions for maxima initiates a variety of methods [1]. The significant part of the methods can be defined as an art of algorithms, some of them build analytical auxiliary functions, some are based on an implementation of well working Taylor series [2].

This variety creates undoubtedly effective tools for global optimization. One can't say however they have mathematical background as the local methods have it in the Taylor series (nearby an objective point $\boldsymbol{u}$):

$$f(x) \approx f(u) + f'(u) \cdot (x-u) + \frac{1}{2} f"(u) \cdot (x-u)^2 + \cdots \quad (1)$$

Moreover the coefficients in (1) are defined in a small vicinity of $\boldsymbol{u}$ (as the derivatives). i.e. ”locally” whereas the searching for global maximum of a multi-extreme function has rather non-local character.

The coefficients of Hermite series (Laurent, Fourier, etc as well) are defined by integrals “observing” entire region of analyzed function ***f(x)*** however such a row doesn’t give a clue what to do then …

The idea was found in symmetrization of a function around objective point $\boldsymbol{u}$:

$$F(x,u) = \frac{1}{2}[f(x) + f(2u-x)] \quad (2)$$

… it’s obviously but very useful:

$$F(u,u) \equiv f(u) \quad (3)$$

The series in weighted Hermite polynomials for function (2):

$$F(x,u) = e^{-2t^2} \sum_{n=0}^{\infty} a_n H_n(t) \quad (4)$$

where

$$t \equiv \frac{x-u}{S(u)\sqrt{2}} \quad (5)$$

$(S(u) is\ defined\ a\ bit\ latter)$

is transformed into a bit promising view (NB (3)!):

$$f(u)\ \approx 1+\mathbf{0}\cdot H_2(0)+\frac{1}{192}\ \boldsymbol{E}(\boldsymbol{u})\cdot H_4(0)+\frac{1}{46080}\ \boldsymbol{P}(\boldsymbol{u})\cdot H_6(0)+\cdots \tag{6}$$

Where:

$$E(u)\equiv-\frac{R^4(u)-2R(u)\cdot\gamma-\ \sigma^2\cdot R_0^2}{S^4(u)} \tag{7}$$

is defined by the parameters integrally characterizing *f(x)*:

$$q\equiv\mu_{o},\ c\equiv\mu_1,\ \sigma^2\equiv\mu_2,\ \gamma\equiv\mu_3,\ \varepsilon\equiv\mu_4/\mu_2^2\text{-}3 \tag{8}$$

$$\mu_n\equiv q^{-1}\int dx\ f(x)(x-c)^n \tag{9}$$

and:

$$R(u)\equiv c-u \tag{10}$$

$$S^2(u)\equiv\ \sigma^2+R^2(u) \tag{11}$$

$$R_o^2\equiv\sigma^2\cdot\ \varepsilon/2 \tag{12}$$

Function (7) plays role of *excess* in Hermite row but to distinguish it from proper excess *ε*(8) of objective function *f(x)* , the name *"Guidance Function"(GF)* will be used in searching for global maximum.

Really the naive example of *f(x)* and *GF* shown in fig.1 demonstrates *GF* as a pointing one to main maximum of *f(x).*

The extreme points of (7) are determined in essential requirement *dE/du=0* that leads to cubic equation for *R(u)*:

$$2\sigma^2R^3+3\gamma R^2+R\cdot R_o^2\sigma^2-\sigma^2\,\gamma=0 \tag{13}$$

This equation has either *one* or *tree* real roots, (14)

that is seen in example of fig.1 The two roots may indicate sub-regions of the biggest maxima. The third one is typical resided at center of function c(8) that means *E(u)* comprises a solution of Mean Least Squares *u=c* that is effective for symmetrical functions (nothing to do more if it is known).

The definition of *P(u)* in (4) is not given… it has similar form as def.(7) of $\sim u^6$. There was a hope this *P(u)* may feel more than 2 big maxima, however, a scrutiny of *P(u)* revealed behavior almost identical to *E(u), P(u)* doesn't see more roots than *E(u)* does.

The coefficient at H_2 in row(4) is zero (that is amazing gift of Dr. Charles Hermite, the Great) due to variable *t*(def.5) playing a key role in a detaching *MLS* from multi-extreme *E(u)*. Thus taking also into account the numerical coefficients in (4) one can write

$$f(u) \sim const + E(u) \tag{15}$$

that can be used as the Guidance Function (*GF(u)*) really in a location of the main maxima positions. Thus the Hermite polynomials gave (15) and disappeared from this task.

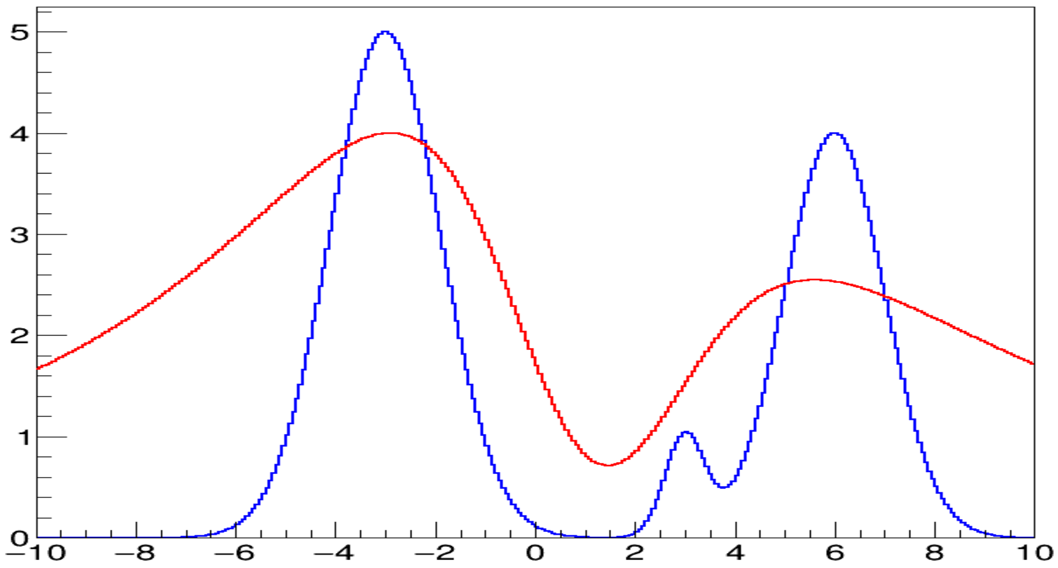

Fig.1 The main maximum of *GF* (in red color) points to the main maxima of objective function.

Convergence

It is seen from previous consideration the *GF*(def.15) provides only three (14) roots. An iterative procedure can be developed for any (unpredictable) function and statement (14) indicates important rule of termination of iterations: if the cubic equation (13) has *one* root only that means *f(x)* is of unimodal character and the *local* method (based on (1)) can be applied.

Thus the number of roots of *GF* plays determinative role in a convergence to global maximum position.

Multi-dimensional Functions

Direct expansion of $\boldsymbol{f}(\vec{\boldsymbol{r}})$ into Hermite polynomials series along each variable x_n of $\vec{r} \equiv \{x_n; n = 1, \dots, N\}$ complicates the task significantly.

One can nevertheless step toward *experimental mathematics* and generalize the definition *GF(u)* for objective vector $\vec{\boldsymbol{u}}$ **:**

$$GF(\vec{u}) = Const + E(\vec{u}) \tag{16}$$

where:

$$E(\vec{u}) \equiv \frac{R^4 - 2\vec{R}\cdot\vec{\gamma} + \vec{R}^T\cdot\hat{K}\cdot\vec{R} + \sigma^2\cdot R_o^2}{(\sigma^2 + R^2)^2} \tag{17}$$

$$\vec{R} \equiv \vec{R}\,(\vec{u}) \equiv\ \vec{c} - \vec{u} \tag{18}$$

$\vec{c}$ is the weighted center of $f(\vec{r})$

$$\sigma^2\ \equiv\ \int f(\vec{r})(\vec{r} - \vec{c})^2\, dv(\vec{r}) \tag{19}$$

$$\vec{\gamma}\ \equiv\ \int f(\vec{r})(\vec{r} - \vec{c})^3\, dv(\vec{r}) \tag{20}$$

$$K_{mn} \equiv 2(\delta_{mn}\sigma^2 - \int f(\vec{r})(x_m - c_m)(x_n - c_n)\, dv(\vec{r})) \tag{21}$$

$$m, n = 1, \dots, N$$

$$2R_o^2\ \equiv\ 3\sigma^2 - \sigma^{-2} \int f(\vec{r})(\vec{r} - \vec{c})^4\, dv(\vec{r}) \tag{22}$$

Illustration in fig.2 confirms possibilities of *experimental mathematics*: *GF(16)* points to main maxima of a test function.

The determination of the positions of the *GF*($\vec{u}$) extrema by $\vec{\nabla} GF(\vec{u})$=0 leads to nonlinear vector equation:

$$Q(\vec{R}) \cdot \vec{R} = \vec{\gamma} -\ \hat{K}\,\vec{R} \tag{23}$$

where

$$Q(\vec{R}) \equiv 2\,\frac{\sigma^2 \cdot R^2 + 2\vec{R} \cdot \vec{\gamma} - \vec{R}^T \hat{K} \overrightarrow{R}\ - \sigma^2 \cdot R_o^2}{\sigma^2 + R^2} \tag{24}$$

There is a way to transform eq. (23,24) to the searching for zeros of ONE dimensional function. It's based on orthogonality of eigenvectors $\overrightarrow{h}$ of symmetrical matrix $\hat{K}$ (21). Expansion of $\vec{R}$ into the eigenvectors row

$$\vec{R}(Q) = \sum_{n=1}^{N} t_n\,(Q) \cdot \vec{h}_n \tag{25}$$

deduces eq.(23) to equation for *t* :

$$t_n(Q) = \frac{\vec{\gamma}\ \cdot \vec{h}_n}{Q + \lambda_n} \tag{26}$$

λ_n is the eigenvalue of $\hat{K}$ (21), and finally the searching for Q-zeros of function:

$$\phi(Q) = Q - Q\left(\vec{R}(Q)\right) \tag{27}$$

may simplify significantly the big dimensional task , just NB (24) has discontinuity of the 1st type at $Q = -\lambda_n$ which is easy resolved.

The example of $\phi(Q)$ is shown in fig.3; the defined Q-zeros allow to find vectors $\vec{R}$ related to extremal points of *GF*($\vec{u}$) that is seen in fig.4.

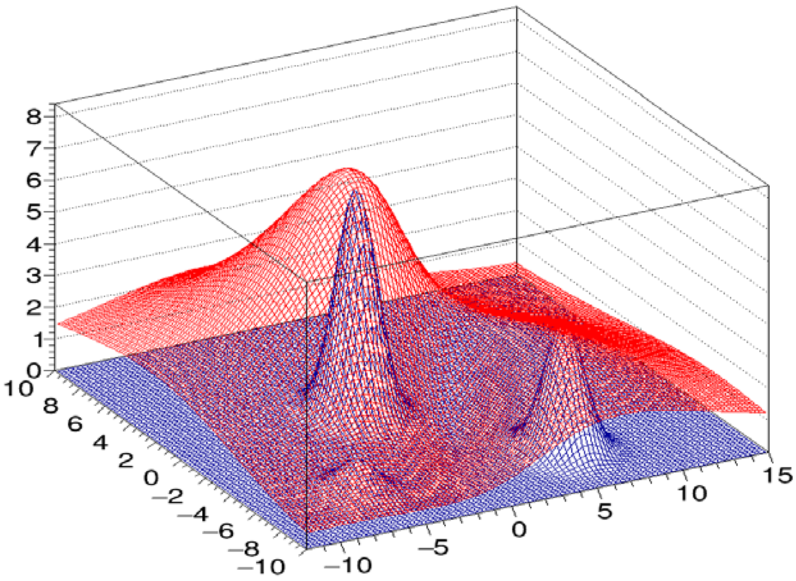

Fig.2 The main maxima of *GF*(def.16) (red) indicate the two big maxima of objective function (shown in blue).

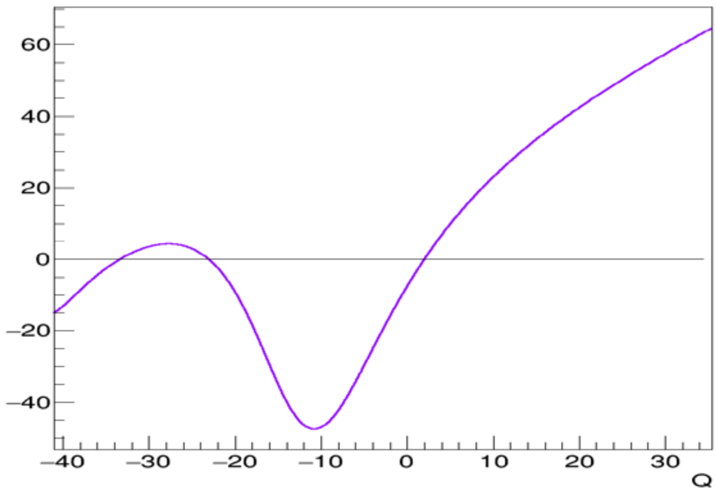

Fig.3 The $\phi(Q)$ (def.27) built for objective function and *GF* shown in fig.2 has the three Q-zeros (($\phi(Q)=0$) .

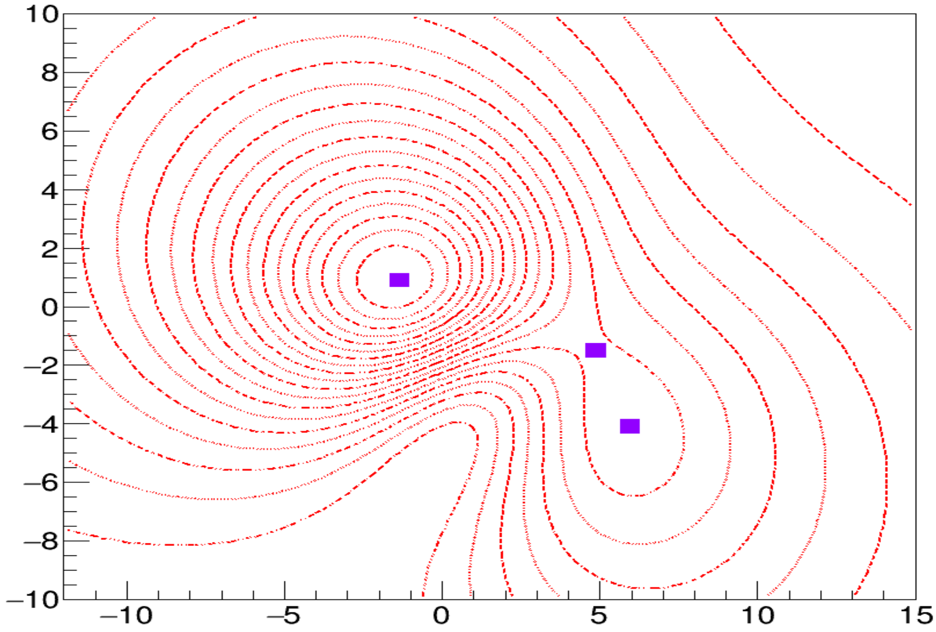

Fig.4 A superposition of *GF* contour (in red) and the three box-points of $\vec{R}$(Q-zeros) (in violet) is shown.

Thus ONE dimensional $\phi(Q)$ substantially simplifies MULTI dimensional searching for global maximum position.

Equalization of Scales

A measure of each "n-th" variable is the dispersion σ_n. The total dispersion σ^2 (def.19) comprises linearly all these σ_n^2, so an unbalance variables (10^6 *and* 10^{-6}) may lead to numerical inaccuracy.

For an avoidance of such problems it is useful to normalize all the variables to any base scale (σ_b):

$$x_n' = \frac{\sigma_b}{\sigma_n}(x_n - c_n)$$

Moreover due to nonlinear character of $GF(\vec{u})$ this re-scaling can improve the guidance, as it seen in fig.5.

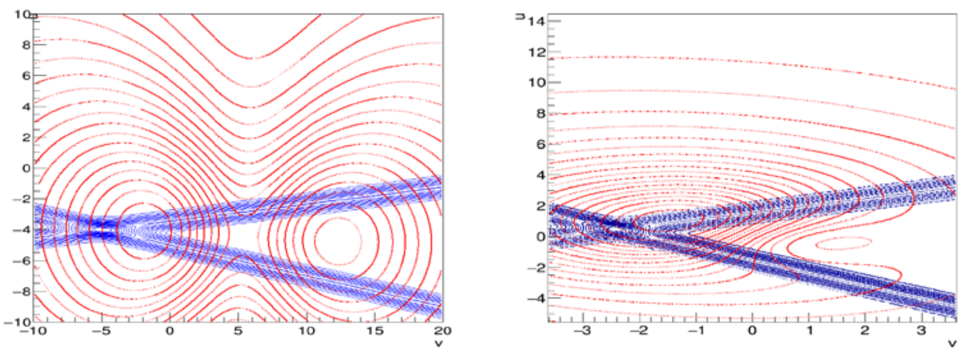

Fig.5 The left plot: a test function (in blue) with 1 main maximum and 2 "mountain range" is recognized by main maximum of *GF* (in red); the *GF*(right plot) for the same *GF* but scaled one has only one maximum and even sees "prongs of mountain rages".

Suppression of a Noise (subtraction of a background)

One can see in fig.5 (left plot) a feature of *GF* : a background, noise imitates fake peak for *GF*. It can be excepted from an incompleteness of (15,16) but level of noise can't be predicted in any precise mathematics.

A nose (background) was added to the test functions (fig. 2) and the *GF* doesn't recognize the global maximum position at all (fig.6)

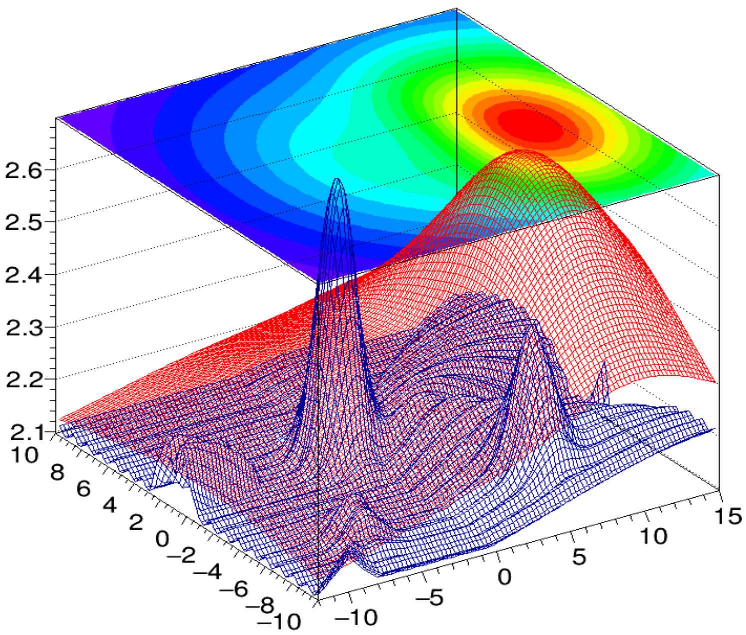

Fig.6 *GF* (in red color, also in contour view on top) points to a vast zone of a noise and doesn't see any big maxima of *f(x)* (blue).

However, the *GF* has noteworthy resource in a suppression of noise.

A presence of a noise can be embodied in analyzed $\phi(\vec{r})$ additively:

$$\phi(\vec{r}) = f(\vec{r}) + B(\vec{r}) \tag{29}$$

where $f(\vec{r})$ is objective function and $B(\vec{r})$ is a background function.

The *integral momenta* (9,19-22) defining *GF* are shown as the "*centered*" ones. These centered momenta can be relatively easy recalculated form *non-centered* ones. Supposition (29) allows to reconstruct non-centered momenta of $f(\vec{r})$

$$\int f(\vec{r}) x_k^m x_l^n \, dv = \int \phi(\vec{r})\, x_k^m x_l^n \, dv - \int B(\vec{r}) x_k^m x_l^n \, dv \tag{30}$$
$$k, l = 1, \dots, N; \;\; 0 \le m + n \le 4$$

The integrals over $\phi(\vec{r})$ represent the *measured* momenta. A known type of the background function may simplify these reconstructions, however, generally, the background type is unknown so it can be a constant *A* in analyzed space V($\vec{r}$).

Thus

$$\mu(f) = \mu(\phi) - A \cdot \mu(1) \tag{31}$$

where

$$\mu(g) \equiv \int g(\vec{r})\, x_k^m x_l^n \, dv \tag{32}$$

The level of noise *A* can be estimated approximately at requirement: all the even non-centered momenta (and derived centered ones) must be > 0.

Fig.7 shows the effect of such approximate estimation of background (rather “noise” in bigger dimension), - the *GF* maxima are very close to the vicinities of the main maxima of objective function.

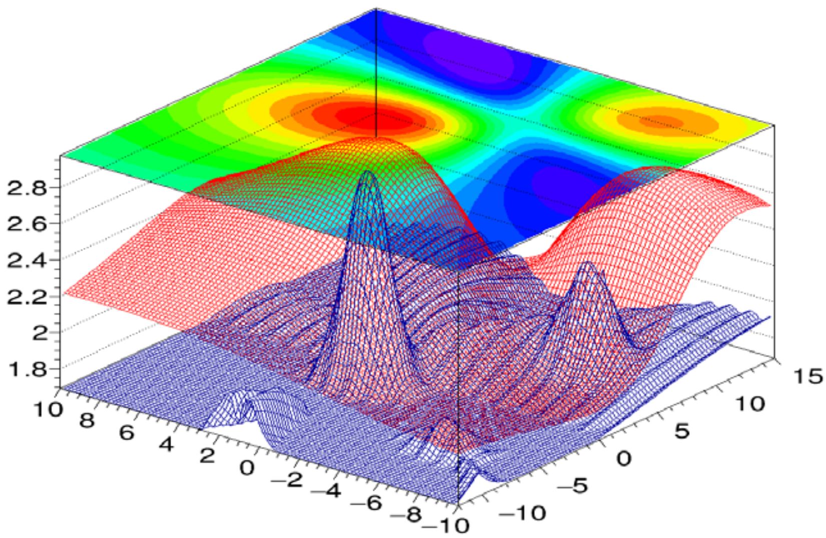

Fig.7 *GF* (in red color, also in contour view on top) points to vicinities of the main maxima due to background subtraction. The objective function (blue) was reduced on the calculated background level $A \approx 0.8$ for demonstration purposes.

Space collapse (Black Hole)

A vast sub-region must get a prior some characteristics like value of dispersion, proper excess etc. A suspected zone can be simply excluded, or be analyzed in details, that increases processing time.

There is an exotic possibility: such zone can be compressed into point (into *nothing*). For a *Black Hole* center placed in $\vec{a}$ at size *D* one can define new variable of *constriction*:

$$\vec{z}(\vec{r}) \equiv \vec{r} + D\,\frac{\vec{r} - \vec{a}}{|\vec{r} - \vec{a}|} \tag{33}$$

The compressed objective function (i.e. $f(\vec{r}) \rightarrow f(\vec{z})$) is deformed significantly nearby *Black*

Hole (D, $\vec{a}$) that is seen in fig.8,9 but the $f(\vec{z})$ *GF* sees already the vicinity of main maximum of $f(\vec{z})$.

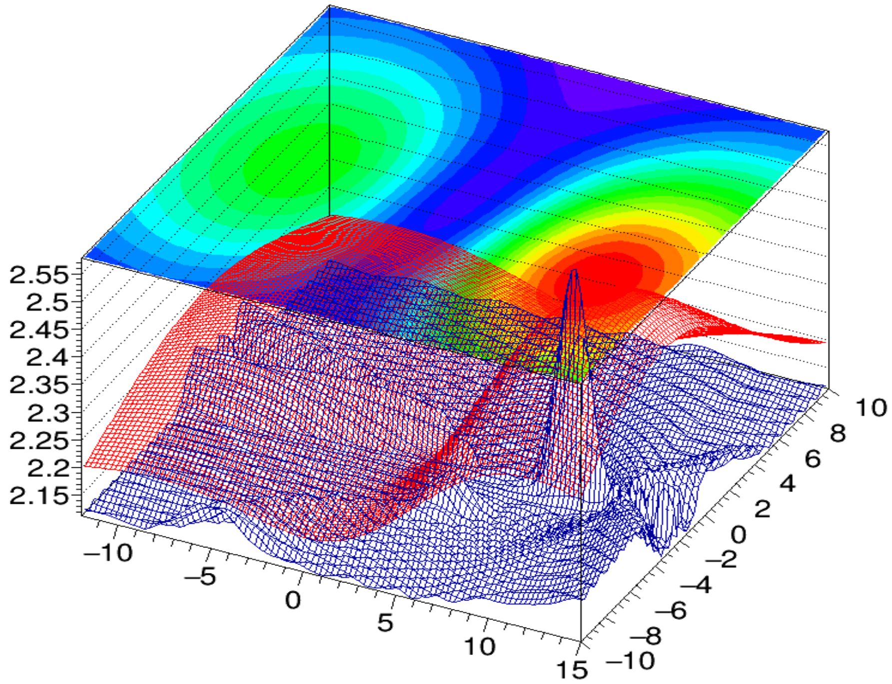

Fig.8 Compressed function saves the main peak on the *Black Hole* edge which absorbed the nose plateau seen in fig.6. The global maximum position of objective function is shifted but in controlled position. *Black Hole* expands small "island" of noise from left into big one so *GF* (in red color, also in contour view on top) marks it. Small parallel ridges demonstrate a deformation of $f(\vec{r})$

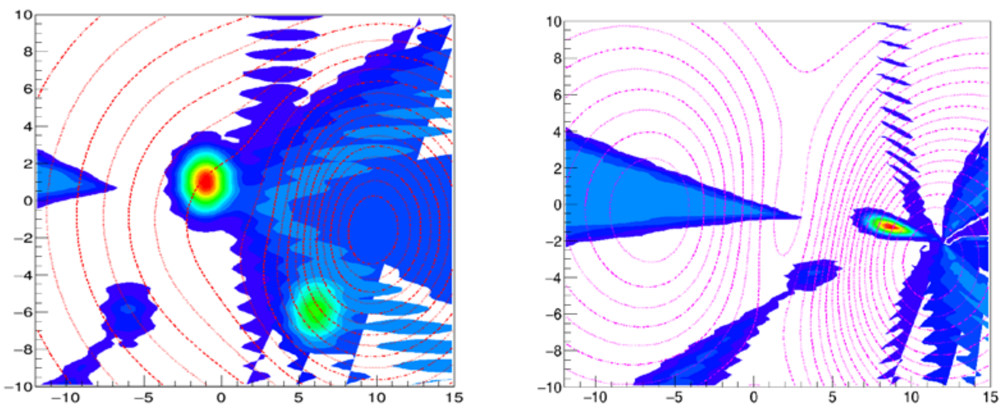

Fig.9 Just a comparison in contour view of scene fig.6 and fig.8 where all the distortions are seen and a fate of the 2^{nd} big maximum $f(\vec{r})$ swallowed by *Black Hole* as well.

Iterations in Real Example

Obviously some efforts should devoted to a definition of Global Maximum: an estimation of minimal level of amplitude, size, shape, etc.

The number of three possible roots of Guidance Function establishes an iterative procedure of a dividing of initial region of $\vec{r}$ into 3(for instance) smaller ones. Such iteration can be repeated until one root is reached at a fulfilment to the Global Maximum requirements. Otherwise this procedure can be continued until entire region will have to be analyzed.

Such iterations have been realized in analysis of experimental data of nuclear interactions in *recursive* form. Fig.10 demonstrates so called *vertex function* [3] determining points of interactions of the ions with the beam pipe. It is good example of multi-extrema function produced by accidentally defocused beam. The peaks of small amplitude (<2) were declared as a noise; the global maximum should have as much associated trajectories as possible (not known exactly, but ≥ 2). The contour view of *vertex* function is shown in fig.11 with superposition of *GF* in the 1st iteration. If a maximum fulfills requirements then iterations are terminated to avoid big recursive depth (huge memory consumption); the pattern is cleared out: the origin of the determinate peak is eliminated, that reduces also a combinatorial background. And iterations begin from initial region. Fig.12 illustrates these steps. The *GF* of most high amplitude took the priority in the iteration queue i.e. “wise iterations”. The background subtraction (31) was applied in this analysis.

Conclusion

The described technologies of analysis of multi-extreme functions may be used in arsenal of Global Optimization. Moreover, the Guidance Function itself can be applied in a prediction for the most probable events in poor statistics.

Acknowledgement

To my daughters Tanya and Nataly for wise advices and support this work, as always.

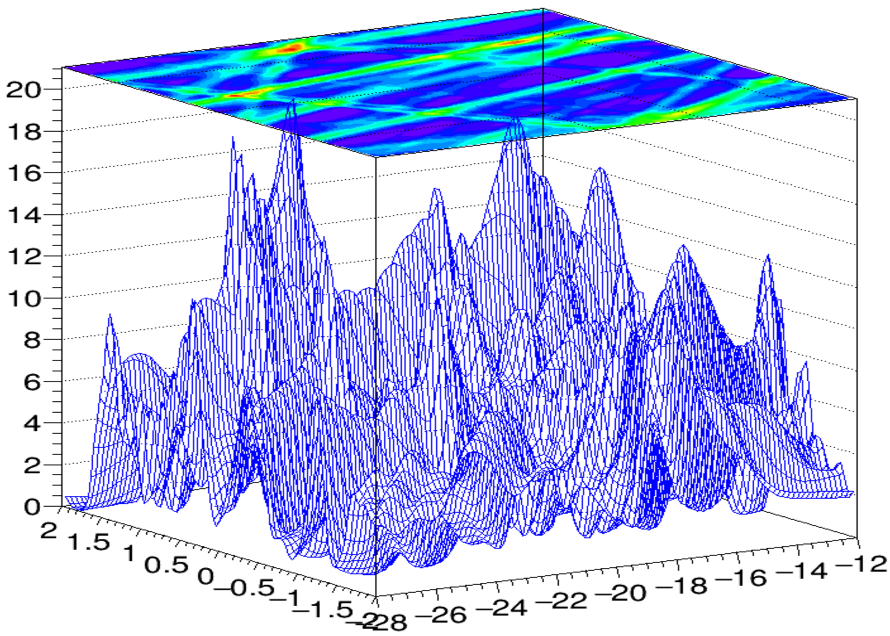

Fig.10 The big maxima of vertex function represent the interaction points of ions on the pipe walls. The background has combinatorial and hardware nature.

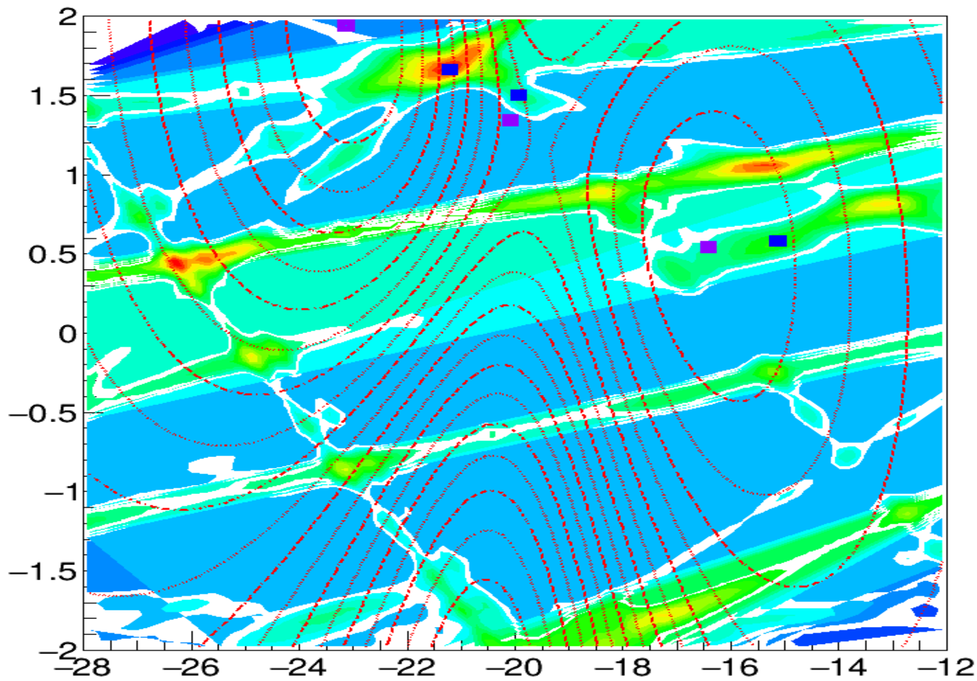

Fig.11 The superposition of contour view of *vertex* function (fig.10) and *GF* (red contour) is shown. The violet squares represents the extremal points of *GF* used then as the start points for *local* searching for maximum (dark blue squares) .

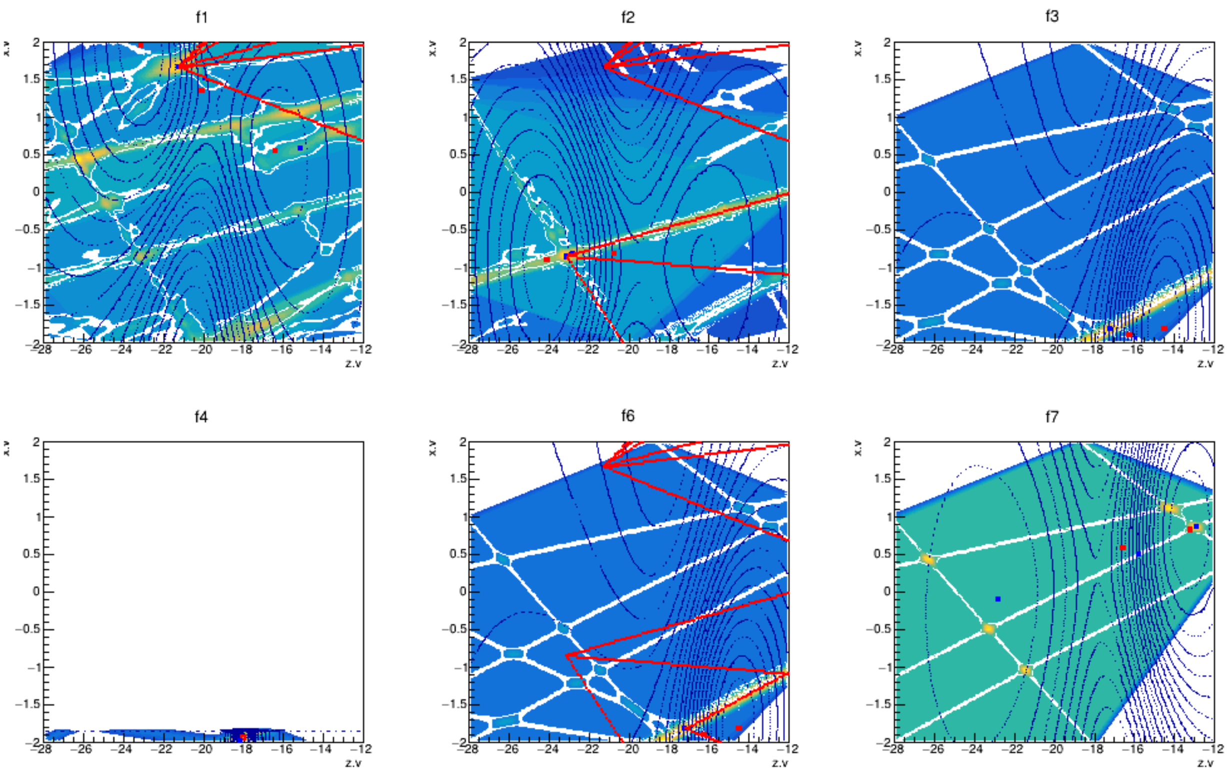

Fig.12 Several recursive iterations of pattern shown in fig.10,11 are demonstrated. The left top picture shows result of the 1st iteration: the global maximum was found which comprises 5 emitted trajectories (shown as red lines) ,… then the 6th iteration gets 2 tracks and analysis has been completed (the noise is only reminded).

References

[1]. http://mathworld.wolfram.com/GlobalOptimization.html
[2]. https://en.wikipedia.org/wiki/Global_optimization
[3]. Ya.D.Sergeev, Journal of Comput.Math. and Math.Phys. 35,No 5,707 (1995)
[4]. A.A. Goldsteai and J.F.Price,Math.Comput.,25,569 (1971)
[5]. Yu.A.Yatsunenko(aka G.A.Vazmin) Nuclear Instruments and Methods In Physics research, NIM A287(1990)422-430