HEURISTIC METHOD FOR SETTING THE INITIAL POINTS IN THE OPTIMISATION PROCESS

J. PUHAN, A. BŰRMEN, T. TUMA, I. FAJFAR, F. BRATKOVIČ
Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia

1. INTRODUCTION

Optimisation methods are used to improve the performance of an analogue electronic circuits[1]. Numerous of papers, describing different algorithms, have been published (see, for example [2, 3, 4, 5 and 6]). The success of the optimisation algorithm greatly depends on the selection of the initial point. The same method can lead to quite different results for variety of initial points. Local methods are usually more sensitive than global ones [7 and 8]. The later have always built in some randomness, which neutralises the importance of the proper selection of algorithms starting point.

The selection of the initial trial is usually left to the user, who relies upon his knowledge and intuition. It is normally chosen in a point, where the circuits best performance is expected. If the estimation is right, the minimum of the cost function lies near and the optimisation task is easy. But on the other hand no additional information is gained. The optimisation process just confirms the expectations. A great part of the parameter space is left unexplored and the question of finding better solution remains open.

Whole parameter space has to be explored to answer the question for sure. So the circuit has to be optimised from several initial points, and each optimisation run has to cover other part of the parameter space. The optimisation process becomes a group of several optimisation runs with this approach. For finding a proper initial trial of each run a probabilistic global method[5] could be employed. Unfortunately it is developed only in one dimension, requires relatively high computational overhead, and is slow in the nearness of the minimum. The paper proposes new heuristic method, which idea basis on probabilistic method[5]. It can be applied in multidimensional parameter space and does not require significant computer effort.

Two artificial mathematical cases illustrate the behaviour of heuristic method. Further the method is applied on two electrotechnic examples: *MOSFET* model parameter extraction and integrated operational amplifier optimisation.

2. ONE DIMENSIONAL APPROACH WITH PROBABILISTIC METHOD

Let cost function $E(\mathbf{x})$, $\mathbf{x} \in A \subseteq \Re^n$, E: $\Re^n \rightarrow \Re$ be real function, where A denotes a feasible region. Purpose of every optimisation process is to find a global minimum $\mathbf{x}_0$ of cost function $E(\mathbf{x})$[1], $E(\mathbf{x}_0) \leq E(\mathbf{x})$, $\forall \mathbf{x} \in A$. In one dimension explicit constraints defines feasible region or parameter space as an interval $A = [x_{low}, x_{high}]$. Let us define a continuous stochastic process $f(x, \omega)$[9]. It assigns a function $f(x)$ to every outcome $\omega \in \Omega$ of an experiment ζ. The domain of ω is the set of all experimental outcomes Ω, and the domain of x is a set of real numbers $\Re$. Let one dimensional cost function $E(x)$ be equal to a realisation of the stochastic process $f(x, \omega)$ for an outcome ω_0 on the interval A.

$$E(x) = f(x, \omega_0) \quad \omega_0 \in \Omega \quad x \in A \tag{1}$$

Cost function $E(x)$ is an arbitrary real function on the interval A. The distribution function $G(f_0, x)$ gives a probability of an event $\{f(x, \omega) \leq f_0\}$ at a particular x by its definition. We presume normal distribution for $G(f_0, x)$[5] with variance $\sigma^2(x)$ and expected value $m(x)$.

$$G(f_0, x) = \mathrm{P}\{f(x, \omega) \leq f_0\} = \frac{1}{\sqrt{2\pi}\sigma(x)} \int_{-\infty}^{f_0} e^{-\frac{(f - m(x))^2}{2\sigma^2(x)}} \, df \tag{2}$$

After one or more optimisation runs the cost function has been evaluated at several points. Lets say we have k such points $x_1, x_2, \ldots x_k$, and the cost function values $E(x_i)$, $i = 1, 2, \ldots k$, are known. Let an event Z_k is defined as a $\{f(x_i, \omega) = E(x_i), i = 1, 2, \ldots k\}$. In other words, the event Z_k occurs, when the stochastic process function $f(x, \omega)$ equals to cost function $E(x)$ in all known points $x_1, x_2, \ldots x_k$, for an outcome ω. The event Z_k becomes certain if expected value $m(x)$ equals to cost function and if variance $\sigma^2(x)$ is zero at all known points. Therefore $m(x_i) =$

$E(x_i)$ and $\sigma^2(x_i) \to 0$ for $i = 1, 2, \ldots k$. When mean and variance has those properties, the distribution $G(f_0, x)$ becomes a conditional probability of an event $\{f(x, \omega) \le f_0 / Z_k\}$.

Let us further set index of a point x_j with the lowest cost function value to *opt*, $x_j = x_{opt}$. So the relation $E(x_{opt}) \le E(x_i)$, $i = 1, 2, \ldots k$, is valid. Further we define a function $f_{min}(x, \omega)$. Its value is always lower than $E(x_{opt})$ for an arbitrary x and any outcome ω.

$$f_{\min}(x, \omega) = \min(E(x_{opt}), f(x, \omega)) \tag{3}$$

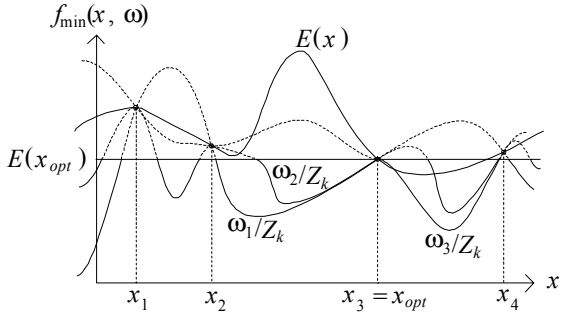

Figure 1: Functions $f_{min}(x, \omega)$ (solid) and realisations of a stochastic process $f(x, \omega)$ (dashed) for different outcomes ω. The event Z_k is certain, therefore $m(x_i) = \mathrm{E}(x_i)$ and $\sigma^2(x_i) \to 0$, $i = 1, 2, \ldots k$, $k = 4$.

Distribution $G_{min}(f_0, x)$ of function $f_{min}(x, \omega)$ gives the probability of an event $\{f_{min}(x, \omega) \le f_0 / Z_k\}$, where Z_k represents a certain event as above. It can be obtained from the distribution $G(f_0, x)$ and the definition of $f_{min}(x, \omega)$. We can also get the density function $g_{min}(f_0, x)$, which is the derivative of distribution $G_{min}(f_0, x)$.

$$G_{\min}(f_0, x) = \mathrm{P}\{f_{\min}(x, \omega) \le f_0\} = G(f_0, x) + (1 - G(f_0, x))u(f_0 - E(x_{opt})) \tag{4}$$

$$g_{\min}(f_0, x) = \frac{\partial G_{\min}(f_0, x)}{\partial f_0} = \frac{e^{-\frac{(f_0 - m(x))^2}{2\sigma^2(x)}}}{\sqrt{2\pi}\sigma(x)}(1 - u(f_0 - E(x_{opt}))) + (1 - G(E(x_{opt}), x))\delta(f_0 - E(x_{opt})) \tag{5}$$

Functions $u(f_0 - E(x_{opt}))$ and $\delta(f_0 - E(x_{opt}))$ in equations (4) and (5) represent a unit step function and its derivative, a unit *Dirac* impulse, respectively.

The expected value $\mathrm{E}\{f_{min}(x, \omega) / Z_k\}$ is a mean of the function $f_{min}(x, \omega)$ at a particular x and different outcomes ω. Because of the event Z_k it equals to the cost functions value $E(x_{opt})$ in all k known points. The question is where to choose new initial point for the next optimisation run, if the cost is already known in k points. A natural decision is to set it there, where the expected value $\mathrm{E}\{f_{min}(x, \omega) / Z_k\}$ is minimal. The lowest cost function value can be expected than, according to our current knowledge. To find out new starting point x_0 a minimisation problem (6) has to be solved. The integral definition of expected value further expresses the minimisation problem with density function $g_{min}(f_0, x)$. Upper bound of definite integral can be set to $E(x_{opt})$ using equation (5).

$$x_0 = \min_{x \in A}(\mathrm{E}\{f_{\min}(x, \omega) / Z_k\}) = \min_{x \in A} \int_{-\infty}^{\infty} f_0 g_{\min}(f_0, x) df_0 = \min_{x \in A} \int_{-\infty}^{E(x_{opt})} f_0 g_{\min}(f_0, x) df_0 \tag{6}$$

The minimisation problem (6) can be transformed into a maximisation problem (7) using the distribution function $G_{min}(f_0, x)$ instead of density.

$$x_0 = \max_{x \in A} \int_{-\infty}^{E(x_{opt})} G_{\min}(f_0, x) df_0 = \max_{x \in A} \frac{\sigma(x)}{\sqrt{2\pi}} \int_{-\infty}^{(E(x_{opt}) - m(x))/\sigma(x)} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt du \tag{7}$$

The distribution and density functions of limited random walk[9], also known as *Wiener* process $w(t)$, are normal with constant mean and variance increasing with t. We also presumed normal distribution for our process $f(x, \omega)$. *Wiener* process $w(t)$ is continuous function of variable t. If we suppose that cost function $E(x)$ is continuous near known points, than it can be a sample path of a *Wiener* process in their nearness. This assumption does not place any physically unrealistic limitations on types of cost functions, which take place in circuit design optimisation problems. Therefore we can presume constant expected value and linearly increasing variance near known points. We set the mean and variance to $m(x) = E(x_i)$ and $\sigma^2(x) = \alpha|x - x_i|$ around i^{th} point. For such setting the event Z_k is certain as well. In nearness of every determined point the equation (7) becomes (8).

$$x_0 = \max_{x \in A} \sqrt{\frac{\alpha|x - x_i|}{2\pi}} \int_{-\infty}^{(E(x_{opt}) - E(x_i))/\sqrt{\alpha|x - x_i|}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt du \quad i = 1,2,\ldots k \tag{8}$$

The expression in equation (8) is monotonically decreasing function of cost value $E(x_i)$ and monotonically increasing function of distance $|x - x_i|$. This leads to two statements. First new initial point x_0 lies rather closer to the known points with lower cost function values, than those with higher ones. And second it lies away form all known points so the distance to the nearest one is large. Both conclusions can be intuitively generalised to n dimensional parameter space. A simple heuristic method, described in continuation, basis on them.

3. A HEURISTIC METHOD FOR FINDING NEW INITIAL POINTS

The second conclusion defines a thesis, that new initial point has to be somewhere in parameter space, where a density of already determined points is low. If it is low, then we expect the average distance between two nearest points to be large in general. But we have to define how to measure known points density. Let us divide the parameter space into 2^n equal subspaces (2^n equal boxes). Let the density be equal to the number of known points in a particular subspace, and let it be constant over whole subspace. New initial point will be chosen in the subspace with the lowest density.

The first conclusion on the other hand tells us, that the contribution to the density is not always the same for all already determined points. Those with lower cost function values should contribute less, than ones with higher cost values. In the previous definition all of them contributed one unit, regardless to the cost function value. Therefore known points have to be weighted. Each point will contribute its weight, which has to be proportional to its cost. Let the weight u of a point with cost function value E be defined with equation (9).

$$u = \frac{(\beta - 1)E + E_{\max} - \beta E_{\min}}{E_{\max} - E_{\min}} \tag{9}$$

E_{min} and E_{max} represent the lowest and the highest cost function value among already determined points, respectively. Point with the lowest value has always weight one. The weight of point with the highest value is given by coefficient β, and now it contributes β times more to the density, than the lowest point.

So far all known points, for which we know, that they violate implicit constraints, are still not included in our definition of density. They do not have cost function value E, so their weight can not be calculated by equation (9). But those points gives us some information about the cost and therefore they have to be taken into account. We set their weight to 2β.

Final heuristic algorithm for determining new initial point for the next optimisation run is described in the repeat until loop below. Space is divided into 2^n equal subspaces, until we find a subspace with no points determined yet. New initial point is selected there randomly. The algorithm is very simple, so it demands only a small amount of computational time.

```
calculate weights for all known points;
temporary space := explicitly constrained space;
repeat
        divide temporary space into 2^n equal subspaces;
        add up weights in particular subspaces;
        temporary space := subspace with the lowest sum of weights;
until lowest sum ≠ 0
randomly pick new point in temporary space;
```

4. NUMERICAL EXAMPLES

4.1 Two mathematical cases

Two different multidimensional and multiminimum cost functions were considered. They are defined with expressions:

$$E(\mathbf{p}) = 1.1 - \sum_{i=1}^{2^n} e^{-r_i^2/2} \tag{10}$$

where

$$r_i^2 = \sum_{j=1}^{n} (p_j - m_j)^2 \quad m_j \in \{2,8\} \quad i = 1,2,\ldots 2^n, \qquad 0 \le p_j \le 10 \quad j = 1,2,\ldots n, \qquad n = 4,$$

and

$$E(\mathbf{p}) = r(3 + \cos 4\varphi + \cos 4\vartheta) \tag{11}$$

where

$$r = \sqrt{p_1^2 + p_2^2 + p_3^2}, \quad \vartheta = \operatorname{arctg} \frac{\sqrt{p_1^2 + p_2^2}}{p_3}, \quad \varphi = \operatorname{arctg} \frac{p_2}{p_1},$$

$$r<1 \vee r>2 \vee (\varphi>\pi/2 \wedge \vartheta<\pi/2), \quad -3 \le p_j \le 3 \quad j=1,2,3.$$

First function (10) has 2^n equal local minima. In our case n = 4, so we have 16 local minima. The second function (11) is defined in three-dimensional parameter space. It has global minimum in the origin, and there are 8 valleys leading to it. All of them are cut through with the implicit constraints, except one.

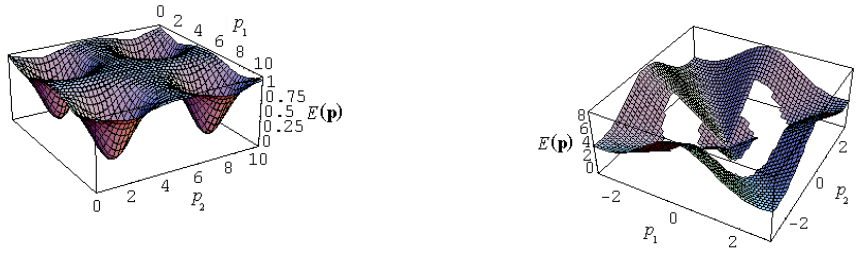

Figure 2: Projection of cost functions (10) and (11) in two-dimensional parameter space

In the first example 16 optimisation runs from initial trials determined with proposed heuristic method were done. In the second example there were optimisation runs performed until the global minimum was reached. We used two gradient (steepest descent and Davidon-Fletcher-Powell[2]) and two direct (Hooke-Jeeves[3] and constrained simplex[4]) methods in optimisation runs.

Table 1: Results for mathematical examples

	first example		*second example*	
method used	*# of different minima found*	*# of cost eval.*	*# of runs*	*# of cost eval.*
steepest desc.	16	651	6	830
DFP	16	3124	6	463
HJ	16	2503	6	974
const. simp.	9	5585	1	219

The results show, that different minima are found in sequential optimisation runs with initial points determined by heuristic. In the first example all 16 local minima were accurately found in 16 runs regardless to optimisation method used. The exception is constrained simplex algorithm which is partially global and is therefore able to climb out of local minimum. But it is slower than other three methods. The second example indicate, that optimisation process travelled through different valley in every run. So sooner or later the right valley to global minimum is found.

4.2 MOSFET model parameter extraction case

In this example we were looking for appropriate combination of parameter values for *SPICE* LEVEL 3 *MOSFET* model. Model parameters were tuned so the $i_{ds}(v_{gs}, v_{bs})$ characteristics would be as close as possible to the measured ones. The cost function was therefore defined as a weighted sum of an absolute differences between measured and modelled drain-source current at different combinations of gate-source and bulk-source voltages[6]. The constrained simplex method was used as an optimisation method in optimisation runs from heuristically determined initial points. Table 2 below summarises the results of optimisation process for *MOSFET* model parameter extraction.

Table 2: Summarised results for the *MOSFET* model parameter extraction

	VTO *[V]*	*U0* *[cm²/Vs]*	*NSUB* *[1/cm³]*	*GAMMA* *[V^1/2]*	*ETA*	*THETA* *[1/V]*	*KAPPA*	*VMAX* *[m/s]*	*# of cost eval.*
heuristic	0.769	901	1.79*e*17	0.928	0.0293	0.998	0.0230	7.70*e*7	2458
initial	0.769	900	1.80*e*17	0.928	0.0293	0.996	1.70	5.26*e*7	2148
*point**	0.769	900	1.81*e*17	0.927	0.0293	0.995	1.34	4.58*e*7	2054
FSD[6]	0.769	900	1.80*e*17	0.928	0.0293	0.996	0.382	5.26*e*7	4258

* typical results (three randomly chosen) of optimisation runs from heuristically determined initial points
fixed model parameters: *XJ* = 200nm, *TOX* = 20nm, *NFS* = 0, *LD* = 100nm, *W* = 10μm, *L* = 1μm

Although the cost function is multiminimum the constrained simplex method did not have any problems with finding a global minimum in a single run from any initial point. The number of cost function evaluations in many trials settles around 2000 evaluations per optimisation run. Comparing to FSD (fast simulated diffusion[6] method based on simulated annealing) approximately two runs from heuristically determined initial points can be performed at the same price. More optimisation runs confirm a global minimum on one side and gives some additional information on the other. In the above example we can see, that all parameters have almost the same values, except *VMAX* and *KAPPA*. Because *VMAX* influences to i_{ds} in a logarithmic manner (increase from 1*e*4

to 1*e*5 tends to have the same effect as increase from 1*e*7 to 1*e*8) we can say that optimal value for parameter *VMAX* is around 5.5*e*7 m/s. So trials from heuristic initial points confirm the optimal values of all parameters except *KAPPA*, which defines channel length modulation in saturation region. Values for *KAPPA* greatly differs among individual runs. That indicate small influence of *KAPPA* to the drain-source current at optimal values of other parameters. This fact can further also be verified by checking the maximum channel length modulation ΔL_{max} versus effective channel length L_{eff}. The ratio $\Delta L_{max}/L_{eff}$ is less than 4e-6, which further means that i_{ds} increases for at most 0.004‰ due to *KAPPA*.

4.3 Integrated operational amplifier optimisation

Let us examine the two-stage operational amplifier[10] shown in figure 3. The optimisation parameters are channel lengths and widths of all transistors and a compensation capacitance C_c. Because pairs of transistors M_1 and M_2, M_3 and M_4, M_7 and M_8 are equal, the parameter space is 11 dimensional.

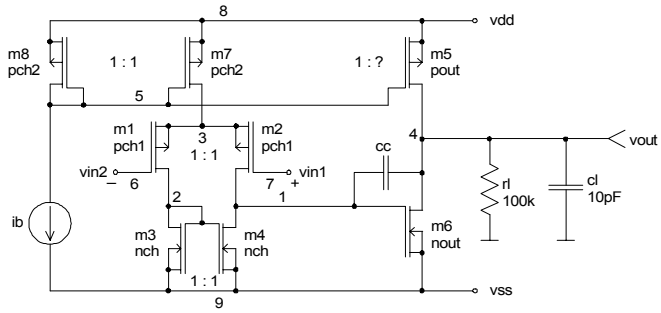

Figure 3: Two stage *CMOS* operational amplifier

When designing an integrated operational amplifier, there are a lot of desired properties to achieve. For example high gain, high bandwidth (f_{-3dB}), high phase margin, small current consumption etc. are wanted. All those requirements are gathered together in cost function. Circuit properties are weighted and their contributions are summed into a final cost value. Therefore it becomes a trade-off among desired properties and the consequence is a multiminimum cost function.

In the first six optimisation runs from heuristically determined initial points we get six different local minima. Three of them are almost equivalent from the cost function point of view. The results are summarised in Table 3. For finding a minimum in a particular optimisation run the constrained simplex method was used. Further the circuit was optimised with genetic algorithm[8]. It finds solutions with relatively low cost value after a reasonable number of cost evaluations. But those circuits have at least one property completely unsatisfied. Less or equal number of cost function evaluations is better used with heuristically determined initial points and greedy optimisation method.

Table 3: Summarised results for the optimisation of the operational amplifier

run	$\|H(j0)\|$*[dB]*	f_{-3dB}*[Hz]*	$\Delta\varphi$*[°]*	*nonlinearity*	U_{pp}*[V]*	I_{tot}*[μA]*	*A[μm²]*	E_{min}	*# of cost eval.*	*total #*
1.	68.2	816	49.4	0.030	3.43	27.1	5297	4.705	2009	
3.	70.5	510	49.7	0.013	3.23	29.2	4328	4.751	2989	15815*
6.	72.1	397	50.6	0.019	3.61	29.8	5720	4.754	2711	
	76.1	129	39.1	0.027	3.44	28.9	1795	6.985	*after* 2281	
GA	70.3	411	39.9	0.013	2.33	18.5	1407	6.424	*after* 4872	22650**
	77.8	175	36.1	0.020	3.70	29.4	1312	5.922	*after* 9040	

* total number of cost function evaluations in six optimisation runs

** no better circuit (with $E_{min} < 6$) found even after total number of cost function evaluations

5. CONCLUSIONS

In this letter a simple heuristic method for setting an initial point is described. Its idea basis on one dimensional probabilistic methods extended to more dimensional parameter space. The method leads to accurate determination of different local minima in several optimisation runs. An arbitrary greedy method could be used in detached run. It turned out that such approach took less or equal evaluations of cost function as some well known global methods (for example: simulated annealing and genetic algorithm), which produces less information (usually only one point, which is probably near global minimum).

The parameter space is generally heavy implicitly constrained in electrotechnical optimisation cases. Finding a feasible initial point is therefore a serious problem. To avoid this, the cost function can be defined as a trade-off among desired properties without any implicit constraints in the first stage of optimisation process. Such cost

function is expected to have lots of minima. For finding them a heuristic method with a greedy optimisation method could be used. Than cost function is changed, so only one or two of the most important circuit properties are included in it. Other properties are given as an implicit constraints and found minima can be used as new initial points in further optimisation process.

REFERENCES

1. R. E. Massara, *Optimisation Methods in Electronic Circuit Design*, Longman Scientific & Technical, Harlow (Essex), 1991
2. R. Fletcher, M. J. D. Powell, *A Rapidly Convergent Descent Method for Minimisation*, Computer Journal, volume 6, pages 163-168, 1963
3. R. Hooke, T. A. Jeeves, *Direct Search Solution of Numerical and Statistical Problems*, Journal of ACM, volume 8, pages 212-229, 1961
4. M. J. Box, *A New Method of Constrained Optimisation and Comparison with Other Methods*, Computer Journal, volume 7, pages 42-52, 1965
5. A. G. Zilinskas, *The One-step Bayesian Method for Seeking the Extremum of the Function of One Variable*, Cybernetics, Vol. 1, pages 139-144, 1975
6. T. Sakurai, B. Lin, A. R. Newton, *Fast Simulated Diffusion: An Optimisation Algorithm for Multiminimum Problems and Its Application to MOSFET Model Parameter Extraction*, IEEE Transactions on Computer-Aided Design, Vol. 11 No. 2, pages 228-234, February 1992
7. J. M. van Laarhoven, E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, Rediel Publishing Company, Dordrecht, 1987
8. D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison Wesley Publishing Company, Reading Massachusetts, 1989
9. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw-Hill Publishing Company, New York, 1991
10. K. R. Laker, W. M. C. Sansen, *Design of Analog Integrated Circuits and Systems*, McGraw-Hill Publishing Company, New York, 1994