On Numerical Approximation of Non-Convex Variational Problems using Stochastic Optimization Algorithm

Aurelija Trgo
Noel J. Walkington
Carnegie Mellon University

Research Report No. 95-NA-004

February 1995

Sponsors

U.S. Army Research Office
Research Triangle Park
NC 27709

National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

On Numerical Approximation of Non-Convex Variational Problems Using Stochastic Optimization Algorithm*

Aurelija Trgo
Noel J. Walkington

Center for Nonlinear Analysis and Department of Mathematics
Carnegie Mellon University, Pittsburgh, PA 15213

November 1994

Abstract

An implementation of the stochastic gradient minimization method is proposed as a viable approach for the solution of non-convex variational problems. In order to numerically characterize the highly oscillatory properties of minimizing sequences of non–convex energies, the approximation of an associated Young measure on a macroscopic mesh is used. However, the presence of a large number of undesired local minima is generic for such problems; consequently, the classical algorithms of descent are inadequate for minimization of these energies. Once a local minimizer is reached, further improvement may be sought through random restart(s). An adaptive filtering is proposed as an ad-hoc step that may be implemented in conjunction with the probabilistic search of feasible directions.

Application of this method to the solution of the variational problem corresponding to the Ericksen–James energy density in two dimensions is demonstrated.

1 Introduction

In this work, a stochastic minimization algorithm is presented for the numerical minimization of non-convex variational problems that arise in phase

*Work supported by the Army Research Office through the Center for Nonlinear Analysis

transitions in solids. The problems of interest involve minimizing energy functionals modeling nonlinear effects of crystal thermoelasticity that lead to difusionless phase transformations. It is well known that, if certain convexity assumptions of the energy density functionals are not satisfied, the infimum of the corresponding total energy will be attained only in some generalized sense. Functions that generate a minimizing sequence of such energies may develop arbitrarily fine oscillations. These oscillations correspond physically to the finely twinned microstructures that can be ubiquitously observed in metallurgy, ferromagnetism etc. The theory of generalized solutions (*Young measures*) introduced by L. C. Young [16] leads to generalized variational problems whose minima characterize such oscillations . Recent refinements of this theory and related developments for various physical problems may be found in [2, 3, 7, 9].

Using the theory of generalized solutions, an elegant numerical scheme for calculating minimizers of non-convex variational problems and their associated Young measures has been proposed by Nicolaides and Walkington [12]. The algorithm itself does not directly compute oscillatory functions; discrete Young measures are used to characterize the properties of highly oscillatory minimizing sequences on scales finer than any computational mesh. For related numerical results, see also [8, 11, 15]. While this algorithm circumvents the resolution problem associated with representing oscillatory functions on a discrete mesh, it is still necessary to minimize a non–convex energy. Computations performed using traditional descent algorithms, such as the gradient or conjugate gradient method, frequently encounter a major drawback: Going immediately downhill as far as possible, such algorithms come to rest at global and local minima indiscriminately. Therefore, in order to address the problem of finding a global extremum in the presence of a large number of undesired local minima, an implementation of the stochastic gradient method is viewed as a viable technique. This approach to the problem enables us to deal with nonconvexity in a more direct way: Once a local minimum is reached, additional improvement may be sought through random restart, in order to jump out of this local minimum and to still be able to overcome upcoming potential barriers (i. e., to "climb hills").

Our method is close in spirit to the method of simulated annealing. Simulated annealing always requires fine tuning of the annealing schedule which dramatically affects the efficiency. To avoid risking that our search effort may be incapable of discovering more promising regions for the optimal solution, certain modifications to the simple simulated annealing algorithm are utilized. More precisely, the notion of *stochastic gradient* is used to

modify random searches; additionally, an auxiliary filter is adopted to average stochastic gradients and to produce more stable estimates of feasible directions.

The algorithm is based on the scheme given in [12], and results are computed for the two dimensional Ericksen-James energy functional. In the following section, the macroscopic formulation of the problem is outlined. The stochastic optimization strategy, together with the ideas behind it, is covered in Section 3.

2 Two-Well Problem and Young Measure Approximation

We consider a simplified formulation of the Ericksen–James variational problem; more detailed accounts may be found in [8, 15]. An outline of the algorithm proposed in [12] for the approximation of the associated generalized solutions is also presented.

Below $\Omega \subset \mathbb{R}^2$ will be a bounded domain, and we shall denote by $u\colon \Omega \to \mathbb{R}^2$ a stable, continuous, and piecewise differentiable deformation of a crystal lattice that has undergone a phase change; the *austenite-martensite* phase transformation in two dimensions. We consider the problem of minimizing the total free energy of the form

$$\mathcal{I}(u) = \int_\Omega F(\nabla u\,(x), \theta)\, dx, \quad u \in \mathcal{W}^{1,p}(\Omega;\, \mathbb{R}^2). \tag{2.1}$$

Here, the energy density F depends on the deformation gradient ∇u and temperature θ. (Below we will assume that the temperature is fixed, and the dependence on the variable θ will be suppressed.) As mentioned in the Introduction, our numerical algorithm will be discussed in the context of the Ericksen–James density given by

$$F(\nabla u) = \Phi(\nabla u^T \nabla u)$$

where,

$$\Phi(C) = \kappa_1\,(tr C - 2)^2 + \kappa_2\, c_{12}^2 + \kappa_3\,[\frac{1}{4}\,(c_{11} - c_{22})^2 - \varepsilon^2]^2, \tag{2.2}$$

with constants κ_1, κ_2, κ_3 and ε derived from physics, and $C = [c_{ij}]$ is a 2×2 matrix. The bulk energy function F satisfies the principle of frame

indifference, and is right invariant under the symmetry group of the lattice. In other words,

$$F(Q\,A) = F(A), \quad Q \in SO\,(2)$$

and

$$F(A\,R) = F(A), \quad R \in \mathcal{G},$$

where $\mathcal{G}$ is the symmetry group of the lattice generated by $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, the rotation of $\pi/2$. At a fixed temperature θ, the functional Φ has an absolute minimum at 2×2 matrices U belonging to the union of two wells that are generated by the action of $SO\,(2)$ on two fixed positive definite symmetric matrices U_0 and U_1 of the form

$$U_0 = \begin{bmatrix} \eta_1 & 0 \\ 0 & \eta_2 \end{bmatrix}, \quad U_1 = \begin{bmatrix} \eta_2 & 0 \\ 0 & \eta_1 \end{bmatrix},$$

where $\eta_1 = \sqrt{1-\varepsilon}$, $\eta_2 = \sqrt{1+\varepsilon}$. When boundary conditions of the form $u\,|_{\partial\Omega} = Bx$ are specified, with

$$B = B_\sigma = \begin{bmatrix} \eta_1\,(1+\sigma\varepsilon) & -\eta_1\sigma\varepsilon \\ \eta_2\sigma\varepsilon & \eta_2\,(1-\sigma\varepsilon) \end{bmatrix}, \quad 0 < \sigma < 1,$$

minimizing sequences $\{u^{(j)}\}$ of (2.1) are known to develop finer and finer oscillations, so it is necessary to admit the generalized solutions of this variational problem. A generalized solution is a pair (u, ν), where $\nu = \{\nu_x\}_{x\in\Omega}$ is a parametrized probability measure, the *Young measure*, that gives the limiting probability distribution of the values of $\{\nabla u^{(j)}\}$ in a vanishingly small neighborhood of each point $x \in \Omega$ as $j \to \infty$ (see, e. g., Ball [2]).

Following [12, 15], we recall that the Young measures is a macroscopic quantity, and can be computed as the solution of the problem

$$\mathcal{I}(u,\nu) = \int_\Omega \langle \nu_x, F(x,\cdot)\rangle\,dx, \quad u \in \mathcal{W}^{1,p}(\Omega), \quad \nu \in \mathcal{YM}(\Omega), \tag{2.3}$$

subject to the constraint

$$\langle \nu_x, I_d\rangle = \nabla u\,(x). \tag{2.4}$$

(The angled brackets denote the expected value of the indicated function with respect to the given probability and I_d is the identity function). In this notation, $\mathcal{YM}(\Omega)$ represents the set of gradient Young measures; that

is, measures derived from sequences of gradients of functions bounded in $\mathcal{W}^{1,p}(\Omega)$.

A numerically subtle issue here is how to approximate a parametrized Young measure $\nu \in \mathcal{YM}(\Omega)$ in the present vector–valued situation. As shown in [9], the Young measure can be approximated by a measure that is piecewise constant in x. Accordingly, the algorithm proposed in [12] uses the standard piecewise linear approximation for the function u on a given triangulation, $\{\mathcal{T}_h\}$ of Ω, and approximations are constructed by minimizing (2.3) over $V_h \times \mathcal{YM}_h$. Here V_h denotes the space of piecewise linear functions on $\{\mathcal{T}_h\}$ and $\mathcal{YM}_h$ the sub-space of Young measures that are piecewise constant on each triangle of $\{\mathcal{T}_h\}$, and are a finite linear combination of Dirac distributions on each such triangle. (Such a combination is dense in the space of probability distributions!) This corresponds to approximating probability distributions by a traditional quadrature rule.

We consider here the simplest case where the discrete measure ν_x consists of precisely two Dirac distributions $\delta(\cdot)$

$$\nu_x = \lambda(x)\,\delta_{A_0(x)} + (1-\lambda(x))\,\delta_{A_1(x)},$$

where $0 \leq \lambda(x) \leq 1$ varies measurably with x. Piecewise constant approximations for A_0, A_1 and λ are considered on $\{\mathcal{T}_h\}$. The algebraic constraint (2.4) now implies that

$$\nabla u = \lambda A_0 + (1-\lambda)\,A_1,$$

so that

$$A_0 = \nabla u - (1-\lambda)\,b \quad \text{and} \quad A_1 = \nabla u + \lambda\,b,$$

with $b = A_1 - A_0$. In order to guarantee that $\nu \in \mathcal{YM}$, a compatibility condition (arising from the fact that the *curl* of a gradient vanishes) requires $A_0 - A_1$ to be a rank one matrix, i.e.,

$$A_1 - A_0 = a \otimes n, \tag{2.5}$$

where $n = (\cos\theta, \sin\theta)^T$ and a is a piecewise constant vector field on $\{\mathcal{T}_h\}$.

In our example, while the matrices U_1 and U_0 do not differ by a rank one matrix, we have

$$Q\,U_1 - U_0 = a \otimes n$$

for some $Q \in SO(2)$. Furthermore, if linear boundary values of the macroscopic deformation u in our problem are considered as above, a minimizing

sequence of (2.1) will develop oscillations, and the generalized solution (u, ν) corresponding to such a sequence will have $u = Bx$ and $\nu = \lambda \delta_{A_0} + (1 - \lambda) \delta_{A_1}$, with A_0 and A_1 satisfying (2.5) and having n parallel to (-1,1). In addition, $\lambda = \sigma$ where $B = B_\sigma$. Hence, our discrete problem reduces to finding parameters $\lambda \in (0,1)$, $a \in \mathbb{R}^2$ and $\theta \in [0, 2\pi)$ that minimize

$$\psi(\lambda, a, \theta) = \lambda F[B_\sigma - (1-\lambda) a \otimes n] + (1-\lambda) F[B_\sigma + \lambda a \otimes n] \to \min \quad (2.6)$$

which has a global minimum energy of zero.

In summary, a generalized solution (u, ν) of the minimization problem (2.3), with the function u satisfying the given boundary conditions, can be exactly represented by our discrete scheme and will produce the global energy minimum. Nevertheless, numerical algorithms for minimizing (2.6) encountered local minima from which the classical descent algorithms could not escape. To eliminate this drawback, an implementation of the method of the stochastic gradient is considered.

3 Stochastic Numerical Optimization

In this section a method of finding feasible descent directions using a stochastic optimization technique, is introduced. We illustrate the operation of this algorithm by minimizing the functional ψ introduced in the previous section with physical constants $\kappa_1 = 10$, $\kappa_2 = 3$, $\kappa_3 = 10$, and the parameters ε and σ are chosen to be 0.1 and 1/3, respectively.

The macroscopic formulation of the problem outlined in the previous section requires minimizing a finite-dimensional non–convex variational problem. Such a problem poses two conflicting goals. On the one hand, a numerical strategy must be capable of learning while searching, gathering global information about the structure of the landscape, and to use it in order to concentrate the search effort in the most promising regions. On the other hand, every such a strategy must also be capable of global exploration, as focusing on the apparently most promising region does not necessarily lead to the discovery of an optimal solution. Therefore, the search strategy must be able to move to various regions of the configuration space and continue searching. In view of these remarks, stochastic algorithms are natural in that they can guarantee extremization of non–convex problems. However, the simple random schemes that search and save the best must be discounted because of efficiency requirements.

A commonly used approach in stochastic numerical optimization is to map an optimization problem into the simulated annealing framework, a procedure that can be highly nontrivial. The following conflicting requirements highlight the practical difficulties encountered with the simulated annealing algorithm (as pointed out in the work undertaken by Gremaud in [6]):

(i) if the temperature is lowered too slowly, the convergence (if any) towards a global minimizer might be too slow to be of any practical use;

(ii) if insufficient time is spent at each temperature (i. e., if the temperature is lowered too abruptly), especially near the freezing point, then the probability of attaining a very low energy configuration is greatly reduced. That is, the solution may get stuck in some local minimum.

Here the "temperature" of the system is a variable that is initially at a high value and is gradually reduced during the search. Since there is always the risk that the system may freeze without having found a global minimum, the selection of an annealing schedule, that is, a decreasing set of temperatures, together with the amount of time to spend at each temperature, requires a substantial amount of art (and, probably, a supercomputer). Up until now, different approaches have been undertaken in order to improve the relative performance of simulated annealing. Ackley [1] proposed an iterated version of the simulated annealing algorithm which restarts if the temperature drops below a minimum threshold. Davis and Ritter [4] have applied a genetic algorithm to determine better annealing schedules. In addition, Gremaud [6] used simulated annealing to construct a numerical test of quasiconvexity.

It is important to recognize that a randomized search does not necessarily imply directionless search. This basically inspired our strategy for solving the given problem: Given an arbitrarily chosen initial condition, some form of a stochastic fluctuation in the direction finding step of the traditional gradient algorithm is implemented. Once a local minimizer is attained in a descent direction, further improvement may be sought through a random restart in our search strategy, in order to "jump" away from such a point. In order to fully test this strategy, our initial data will be chosen as the *local minimum* of ψ given by

$$x^{(0)} = [\lambda^{(0)}, a_1^{(0)}, a_2^{(0)}, \theta^{(0)}]^T,$$

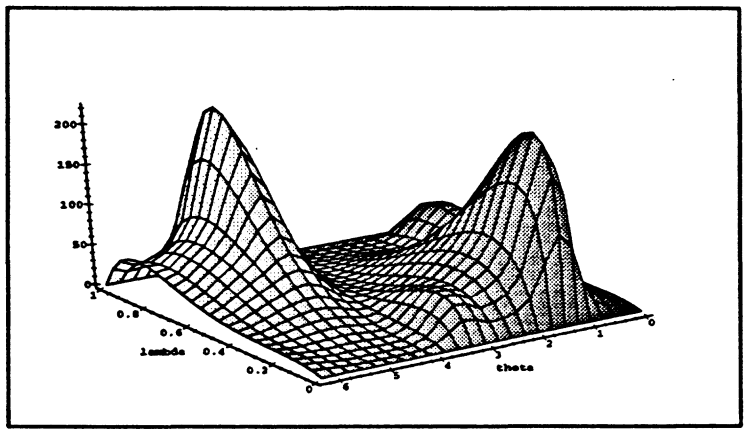

Figure 1: *Level surface of λ and θ components of ψ for $a_1^{(0)}$ and $a_2^{(0)}$. Algorithm started at the local minimum $x^{(0)}$.*

obtained by a routine Maple calculation and having the values

$$\begin{aligned} \lambda^{(0)} &\approx 0.99786045 \\ a_1^{(0)} &\approx 0.69792791 \\ a_2^{(0)} &\approx -1.9104128 \\ \theta^{(0)} &\approx 1.87752784, \end{aligned}$$

The corresponding local energy minimum is $\psi^{(0)} \approx 0.0006$. It is *a priori* known that the correct solution of (2.6) has $\lambda = 1/3$, $a \approx [0.1342, 0.1483]^\tau$ and $\theta = 3\pi/4$, and for these values of the parameters a global energy minimum of zero is attained. Figure 1 plots a 3D level surface of the the functional ψ, with $a_1^{(0)}$ and $a_2^{(0)}$ fixed. This plot shows that ψ is very flat in the region near the point $x^{(0)}$. This could render the algorithm quite "dependent" on the number of correct decimal places used for computation.

3.1 Stochastic gradient algorithm

We begin here with a simplified formulation of the *stochastic gradient method* for minimizing a nonconvex function $\phi : \mathbb{R}^d \to \mathbb{R}$. The method is defined in following terms: Given initial condition $x^{(0)}$, we set

$$x^{(k+1)} = x^{(k)} - h_k\,(x^{(k)})\,\xi^{(k)}, \quad k = 0, 1, \ldots, \tag{3.1}$$

where, at every iteration step k, a sequence of random vectors $\xi^{(k)} = \xi(x^{(k)})$ called *stochastic gradients* corresponding to $\nabla\psi(x^{(k)})$, is generated, having expectation $\mathbb{E}(\xi^{(k)}|x^{(k)})$ equal to a gradient of ψ at $x^{(k)}$. Additionally, $h_k = h_k(x^{(k)})$ is the stepsize at the $k-$th iteration. We remark that, for the sake of numerical simplicity, all probabilistic characteristics of the vector $\xi^{(k)}$ are assumed to be fully determined by the point $x^{(k)}$ and do not depend on the past of the search process. It is possible to show, using stochastic analysis techniques, that such a sequence of iterates will tend to a global minimizer of ϕ "in some probabilistic sense", provided that the sequence $\{h_k\}$ satisfies certain summation properties; namely,

$$\sum_{k=0}^{\infty} h_k = \infty, \quad \sum_{k=0}^{\infty} h_k^2 < \infty, \tag{3.2}$$

and the sequence $\{\mathbb{E}(\| \xi^{(k)} \|^2 \mid x^{(k)})\}_{k=0}^{\infty}$ is uniformly bounded. This result, and the more details, can be found in, e. g., [5, 14]. However, conditions (3.2) are not necessary to guarantee the corresponding computational scheme to converge efficiently to a global minimum (take, for example, the harmonic sequence $\{h_k = 1/(k+1)\}$). Therefore, a direct implementation of such an algorithm can result in a numerically inadequate search strategy.

We use a modification of the basic approach to (3.1) by introducing an auxiliary sequence of *filters*, $z^{(k)}$ which "average" stochastic gradients to produce more stable estimates of the feasible directions that are, in turn, used to compute $x^{(k)}$. To be more precise, we first set

$$\xi^{(k)} = g^{(k)} + \alpha_k r^{(k)},$$

where $g^{(k)} = \nabla\psi(x^{(k)})$ and $r^{(k)}$ is a standard Gaussian randomly generated vector. A control parameter α_k is used in order to weigh the random deviations in the components of $r^{(k)}$, which are in turn found to be quite sensitive to the way random numbers are generated. To this end, the filtering is introduced in the following way:

$$\begin{cases} z^{(k+1)} &= z^{(k)} + \tilde{h}_k (\xi^{(k)} - z^{(k)}) \\ z^{(0)} &= 0. \end{cases} \tag{3.3}$$

Now (3.1) becomes

$$x^{(k+1)} = x^{(k)} - \tilde{h}_k z^{(k+1)}, \quad k = 0, 1, \ldots,$$

The parameter $\tilde{h}_k$ plays the role of the diffusion coefficient, and is decreased at an experimentally determined rate that is analogous to tuning the annealing rate with simulated annealing (see, e.g., [10]). In fact, the appropriate choice of $\tilde{h}_k$ of the form

$$\tilde{h}_k = \frac{c}{\sqrt{\ln(k+2)}}, \tag{3.4}$$

with $c \in (0,1)$, turned out to lead to convergence of our algorithm, as shown below.

This algorithm has no explicit switch for detecting when it is at a local minimum. Instead of actually detecting local minima and restarting, it primarily relies on its ability to accept uphill moves to escape local minima. Although the uphill random moves are as likely to be accepted as those that are downhill, the numerical implementation of the filtering procedure (3.3) was shown, in turn, to provide a greater chance for the moves that were supposed to improve the search.

The method we propose here has close connections with the work of Ruszczyński [13], where related analytical issues may be found. Such algorithms can be adapted to more general settings; for example, to the method of stochastic *sub*gradients (or $\varepsilon-$subgradients) which arises in nonsmooth optimization (see e. g., [14]).

3.2 Numerical Results

This modified stochastic gradient algorithm with filtering is found to be moderately effective for computating the optimal solution to the problem (2.6). After 80 Maple iterations, each requiring only a single function evaluation of ψ, the algorithm gets close to the global minimum. Indeed, it gives a very satisfactory answer of the form

$$\begin{aligned} \lambda^{(1)} &\approx 0.4017524 \\ a_1^{(1)} &\approx 0.13873311 \\ a_2^{(1)} &\approx 0.14478447 \\ \theta^{(1)} &\approx 2.38171172, \end{aligned}$$

when we use $c = 0.3$. The associated energy is $\psi^{(1)} \approx 0.000725$. A 3D level surface plot of ψ, corresponding to $a_1^{(1)}$ and $a_2^{(1)}$ fixed as above is exhibited in Figure 2 below.

As can be clearly observed from this plot, the region around the global minimum is small and is contained in a very narrow and long valley between

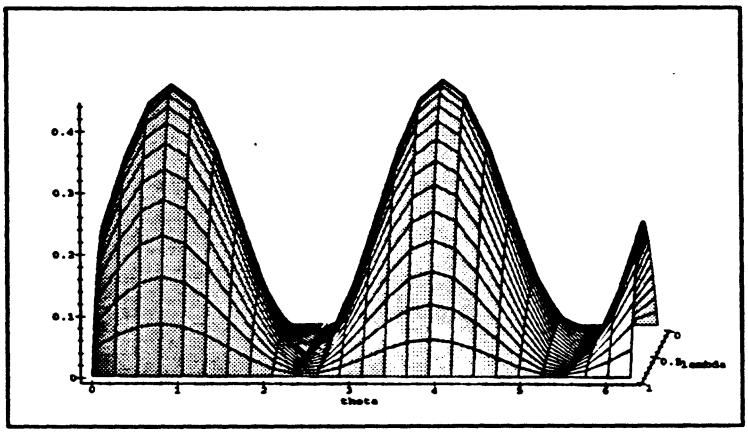

Figure 2: *Level surface of λ and θ components of ψ for $a_1^{(1)}$ and $a_2^{(1)}$.*

two high peaks. In fact, the gradient $\nabla\psi$ at the point $\boldsymbol{x}^{(1)} = [\lambda^{(1)}, a_1^{(1)}, a_2^{(1)}, \theta^{(1)}]^T$ is

$$\nabla\psi(\boldsymbol{x}^{(1)}) \approx [0.01015, 0.118804, -0.096992, -0.023771]^T.$$

The functional ψ is locally convex in the neighborhood of the global minimizer; therefore, the stochastic fluctuations cause rapid variations in the gradient directions as the algorithm gets closer and closer to the global minimum. Hence, in order to gain further improvement in minimizing ψ at this point, additional "numerical trickery" is applied. In fact, after just 9 iterations of the simple downhill descent via conjugate gradients, a solution of

$$\begin{aligned} \lambda^{(2)} &\approx 0.39618443 \\ a_1^{(2)} &\approx 0.13175634 \\ a_2^{(2)} &\approx 0.14201515 \\ \theta^{(2)} &\approx 2.3568941, \end{aligned}$$

gives $\psi^{(2)} \approx 0.000063$, before the algorithm starts leading to an underflow in the computation of the gradient.

We remark here that the numerical performance of the proposed algorithm has been considered by repeating the computations with various different values of physical constants κ_1, κ_2, κ_3, ε and σ, as well as varying the annealing coefficient c in (3.4). As an overall comparison, the efficiency of the procedure does not change significantly from case to case, requiring

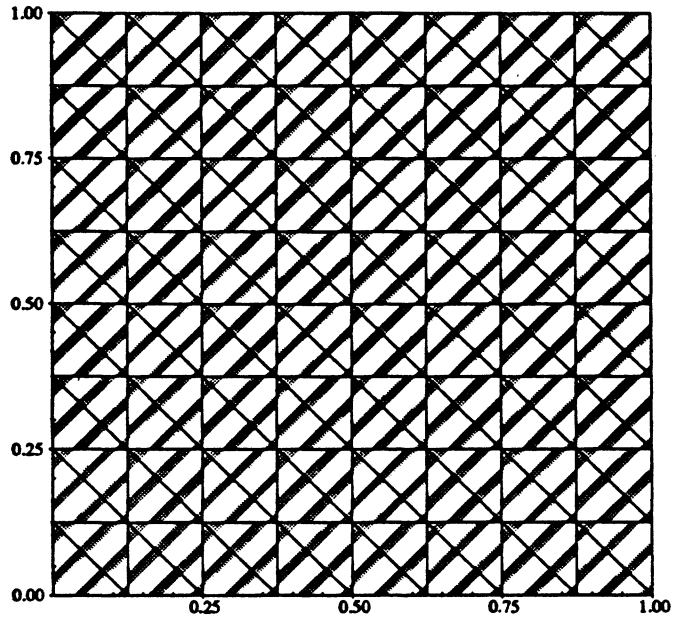

Figure 3: *Rotationally Invariant Wells, 8x8 Mesh.*

about 60 – 120 Maple iterations with the noise terms, before the algorithm gets close enough to the global minimizer. Figure 3 shows a layered microstructure corresponding to the Young Measure computed here having a ratio of widths $\lambda = 1/3$, and layer normals $\boldsymbol{n}$ are at 135 degrees from the horizontal.

3.3 Concluding Remarks

In conclusion, the overall approach is a useful one: The algorithm allows us to start from a local minimum as the initial condition without getting stuck there, in order to get to a well where a global minimum lies, contrary to traditional descent methods. Nevertheless, the performance of these algorithms is very problem dependent.

Moreover, a subtlety associated with the proposed algorithm is the fact that it is very sensitive to the way random numbers are generated. The computations presented above have been performed using a DEC 5000/25 work station, in conjunction with Maple's built-in random number generator. Some interesting but undeveloped ideas involving the use of *genetic algorithms* [1] to improve such randomized searches and implement them into a more general case are currently under consideration.

Acknowledgements. We wish to express our sincere thanks to S. SHREVE for some valuable suggestions. Also, we thank A. DE SIMONE and G. FRIESECKE for enlightening conversations related to the problem.

References

[1] D. H. ACKLEY, *An Empirical Study of Bit Vector Function Optimization,* in Genetic Algorithms and Simulated Annealing, L. Davis, ed., (1987), pp. 170-204.

[2] J. M. BALL, *A version of the fundamental theorem for Young measures,* in Partial Differential Equations and Continuum Models of Phase Transitions, D. Serre, ed., Springer, May 1988.

[3] J. M. BALL AND R. D. JAMES, *Fine phase mixtures as minimizers of energy,* Archive for Rational Mechanics and Analysis, 100 (1987), pp. 13-52.

[4] L. DAVIS AND F. RITTER, *Schedule optimization with probabilistic search,* Proceedings of the Third IEEE Conference on Artificial Intelligence Applications (1987).

[5] Y. N. ERMOL'EV AND N. Z. SHOR, *A random search method for two-stage problems of stochastic programming and its generalization,* Kibernetika (Kiev), No. 1, (1968), pp. 90-92.

[6] P. A. GREMAUD, *Numerical optimization and quasiconvexity,* IMA Tech. Report No. 1133, April 1993.

[7] R. D. JAMES AND D. KINDERLEHRER, *Theory of diffusionless phase transitions,* PDE's and Continuum Models of Phase Transitions, Lecture Notes in Physics, 344, Springer 1989, pp. 51-84.

[8] D. KINDERLEHRER, R. A. NICOLAIDES AND H. WANG, *Spurious oscillations in computing microstructures,* CNA Tech. Report No. 93-NA-009, February 1993.

[9] D. KINDERLEHRER AND P. PEDREGAL, *Weak convergence of integrands and the Young measure representation,* The Journal of Geometric Analysis, 4 (1) (1994), pp. 59-90.

[10] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations,* Springer-Verlag, (1992), pp. 274.

[11] L. MA AND N. J. WALKINGTON, *On algorithms for non-convex optimization,* SIAM Journal on Numerical Analysis, (1993) (in print).

[12] R. A. NICOLAIDES AND N. J. WALKINGTON, *Computation of microstructure utilizing Young measure representations,* J. Intelligent Material Systems and Structures, 4 (1993), pp. 457-462.

[13] A. RUSZCZYŃSKI, *A method of feasible directions for solving nonsmooth stochastic programming problems,* Lecture Notes in Control and Information Sciences, 76, Springer 1983, pp. 258-271.

[14] N. Z. SHOR, *Minimization Methods for Non-Differentiable Functions,* Springer, 1985.

[15] N. J. WALKINGTON, *Numerical approximation of non-convex variational problems,* Smart Structures and Materials, Vol. 2192, 1994.

[16] L. C. YOUNG, *Lectures on the Calculus of Variations and Optimal Control,* Chelsa, 1980.