Variational principle in problems involving instantons

D. I. D'yakonov and V. Yu. Petrov

B. P. Konstantinov Leningrad Institute of Nuclear Physics, Academy of Sciences of the USSR
Submitted 1 January 1982
Zh. Eksp. Teor. Fiz. **86**, 25–38 (January 1984)

A simple approximate method is proposed for the evaluation of path integrals in problems in which subbarrier transitions (instantons) play an important role. The method is based on the Feynman variational principle, and is a generalization of the quasiclassical approximation. When the problem involves a quasiclassical parameter, the method reduces to the approximate evaluation of Gaussian functional integrals (i.e., functional determinants). For such cases, general formulas are derived and are suitable for a wide class of potentials. The accuracy of the approximation is illustrated by two-well and periodic potentials. In both cases, the level splitting can be determined to within a few percent.

1. INTRODUCTION

The quasiclassical method of evaluating functional integrals is frequently used in field theory. It is valid when physical considerations show that the principal contribution to a functional integral is due to a particular class of field configurations, and the fluctuations around them are small. The quasiclassical approach then assumes that the selected field configuration around which small fluctuations take place will satisfy the equation of motion. On the other hand, a typical situation is that where field configurations that are natural from the physical point of view do not satisfy the equation of motion. For example, in quantum chromodynamics (QCD), there are good reasons to suppose that instanton-anti-instanton configurations[1,2] play an important role in vacuum field fluctuations, and the number of instantons should be equal to the number of anti-instantons to within thermodynamic fluctuations $\sim\sqrt{N}$ since, otherwise, CP-invariance will be strongly violated. Although the instanton and the antiinstanton will separately satisfy the QCD equation of motion, an arbitrary superposition of these solutions is not a solution of the equation of motion. Analogous examples can be found in other theories, too (see, for example, Ref. 3).

We note that configurations that do not satisfy the equation of motion can, in principle, provide a greater contribution to the functional integral (partition function) than classical trajectories that ensure local minimum of action, for the simple reason that they have greater entropy (statistical weight). Moreover, if we use the language of statistical physics, we are always interested in the minimum of free energy rather than energy (action).

Thus, we have the relatively general problem of evaluating the contribution to the functional integral of field configurations that, generally speaking, are not solutions of the equations of motion, together with small fluctuations around them. It seems to us that a convenient method that will enable us to solve this problem, and will provide us with a generalization of the quasiclassical theory, is to us the Feynman variational principle.[4]

Let us briefly recall the principle of this method. Suppose that we wish to evaluate the functional integral over the fields

$$Z=\int D\varphi e^{-S[\varphi]}, \tag{1}$$

where the action $S[\varphi]$ is too complicated for precise evaluation. Let us therefore replace the exact $S[\varphi]$ with the approximate action $S_1[\varphi]$ that satisfies the following conditions:

1) on a trial field configuration $\bar{\varphi}$, which we expect to provide the main contribution to (1), the approximate action is either identical with or close to the exact action

2) $S_1[\varphi]$ "clamps down" the field fluctuations around the trial configuration $\bar{\varphi}$. In other words, the integral of $\exp(-S_1)$ must be constructed so that the chosen trial configuration does, in fact, provide the main contribution to it

3) $S_1[\varphi]$ must be sufficiently simple to enable us to evaluate the approximate partition function

$$Z_1=\int D\varphi e^{-S_1[\varphi]}. \tag{2}$$

The last two requirements mean, in practice, that S_1 must be a positive-definite quadratic form in the deviation $\varphi-\bar{\varphi}$.

Let us rewrite the exact partition function (1) in the form

$$Z=\frac{\int D\varphi\exp\{-(S[\varphi]-S_1[\varphi])\}\exp\{-S_1[\varphi]\}}{\int D\varphi\exp\{-S_1[\varphi]\}}\times\int D\varphi\exp\{-S_1[\varphi]\}$$

and use the convexity condition

$$\langle e^{-x}\rangle\geqslant e^{-\langle x\rangle}.$$

This provides us with the lower bound for the exact partition function:

$$Z\geqslant Z_1\exp\{-\langle S-S_1\rangle\}, \tag{3}$$

where

$$\langle S-S_1\rangle=\frac{1}{Z_1}\int D\varphi(S[\varphi]-S_1[\varphi])e^{-S_1[\varphi]}. \tag{4}$$

It is clear that the exact Z does not depend on the trial configuration $\bar{\varphi}$, whereas the approximate expression for Z, given by the right-hand side of the inequality (3), is a function of $\bar{\varphi}$. Hence, by varying the right-hand side of (3) with respect to $\bar{\varphi}$, we can find the maximum (3) that corresponds to the best approximation. If the approximate action S_1 depends on a number of parameters or functions, the variation must also be performed with respect to these quantities if we are to obtain the best approximation.

Relation (3) is actually the first term of a series for Z. The accuracy can be improved somewhat by using the following formula for small $S - S_1$:

$$Z \approx Z_1 \exp(-\langle S-S_1\rangle) \exp\{{}^1/_2[\langle (S-S_1)^2\rangle - \langle S-S_1\rangle^2]\}. \quad (5)$$

If, on the other hand, $S - S_1$ is not small, this formula enables us to estimate the error of a given calculation. We emphasize that, even when this error is large as a result of an inadequate choice of S_1, the inequality (3) may still be useful in that it will provide us with an estimate for the lower bound for the exact Z.

We note that the usual procedure is to evaluate not the partition function Z itself, but a quantity divided by the partition function with the free action $S_0[\varphi]$. This does not invalidate the variational principle (3); we need only remember that Z_1 is also divided by the partition function. We also note that, in the renormalized field theory, the quantity Z does not exist without regularization at high field frequencies. To preserve the inequality given by (3), one must then explicitly introduce regularization (in fact, the same regularization) for both the exact Z and the approximate Z_1, for example, dimensional, lattice, and so on, regularization.

We emphasize that the variational principle given by (3) is a generalization of the quasiclassical theory in the sense that if, as a result of the variation (3) with respect to the trial configuration $\bar{\varphi}$, it turns out that $\bar{\varphi}$ is a solution of the equation of motion, the variational principle will yield the same answer as the usual quasiclassical theory. We shall demonstrate this important point by considering an elementary example involving the evaluation of a one-dimensional integral.

Suppose we have to evaluate the integral

$$Z = \int_{-\infty}^{+\infty} dx e^{-f(x)},$$

and suppose also that we suspect that the main contribution to the integral is provided by the neighborhood of some point $\bar{x}$. We take the approximate "action" to be $f_1(x)$, which satisfies conditions (1)–(3) and is given by

$$f_1(x) = f(\bar{x}) + {}^1/_2 a^2 (x-\bar{x})^2,$$

where a is the variational parameter. We emphasize that, although $\bar{x}$ is not *a priori* a "solution of the equation of motion" $f'(\bar{x}) = 0$, the above logic, nevertheless, forces us to take the approximate "action" $f_1(x)$ without the linear term $f'(\bar{x})(x - \bar{x})$. In fact, unless this is done, integration over $x - \bar{x}$ would involve a translation and, essentially, we would be evaluating the contribution not of the region around the trial point $\bar{x}$ but around the true saddle in the original integral. On the other hand, it is precisely the contribution of the region around the more or less arbitrary point $\bar{x}$ to Z that is of interest to us here.

Let us evaluate

$$Z_1 = \int dx e^{-f_1(x)} = \frac{\sqrt{2\pi}}{a} e^{-f(\bar{x})},$$

and

$$\langle f-f_1\rangle = \frac{1}{Z_1}\int dx e^{-f_1(x)}[f(x) - f_1(x)]$$

$$= \frac{1}{Z_1}\int dx e^{-f_1(x)}\Big[f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + \frac{f''(\bar{x})}{2}(x-\bar{x})^2 + \ldots - f(\bar{x}) - \frac{a^2}{2}(x-\bar{x})^2\Big] \approx \frac{f''(\bar{x}) - a^2}{2a^2}.$$

This gives us the lower bound for the original integral Z:

$$Z \geqslant Z_1 \exp\{-\langle f-\bar{f}\rangle\} \approx \frac{\sqrt{2\pi}}{a} e^{-f(\bar{x})} \exp\Big\{-\frac{f''(\bar{x}) - a^2}{2a^2}\Big\}.$$

By varying this expression with respect to the parameter a, we obtain the best $a^2 = f''(\bar{x})$ (as expected). The final result, therefore, is

$$Z \geqslant \Big[\frac{2\pi}{f''(\bar{x})}\Big]^{1/2} e^{-f(\bar{x})}.$$

We must now vary this expression with respect to $\bar{x}$ in order to determine the $\bar{x}$ whose neighborhood provides the main contribution to the required integral. It is clear that, if $f(x)$ is a "sharp" function, the best $\bar{x}$ is determined by the solution of the "classical equation of motion" $f'(\bar{x}) = 0$, and the variational principle leads us to the same answer as the saddle-point method.

When the number of degrees of freedom is infinite, the variational principle again does not allow the possibility that the trial configuration satisfies the equation of motion, but it does allow us to estimate the contribution to the partition function of more general configurations. We emphasize once again that it is not the action (or the energy, in the language of statistical physics), but the free energy, that must be minimized, i.e., the entropy of the configurations must be taken into account. In contrast to quasiclassical theory, the variational principle is, in fact, suitable for the evaluation of the minimum of free energy.

We note one further advantage of the variational method as compared with standard quasiclassical theory. The basic technical difficulty in the latter theory is the evaluation of the Gaussian integral over small deviations around the chosen classical motion (functional determinant). As a rule, this difficult evaluation is also complicated by the fact that the zero modes have to be isolated, and integration over them must be performed separately. The variational method, on the other hand, provides us with a very effective procedure for the approximate evaluation of functional determinants, even for external fields $\bar{\varphi}$, for which the exact evaluation of the Gaussian integrals is a hopeless task. We consider that the development of this procedure is the principal result reported in the present paper.

Our aim in this research was to demonstrate the application of the above variational principle to quantum-mechanical problems in which instanton-anti-instanton subbarrier transitions play an essential role. From the point of view of path integrals, quantum mechanics may be looked upon as a one-dimensional quantum field theory. The generalization of the method to the four-dimensional theory, i.e., quantum chromodynamics, will be considered elsewhere.

More specifically, we consider two problems, namely, a two-well potential with a high barrier, and a periodic potential of the form $\cos x$. In the former case, we are interested in the splitting of low-lying levels and, in the second, in the width of the first forbidden band. In the limit of high barriers, the fact that the principle contribution to the path integral in these problems gives a tenuous instanton-anti-instanton gas is well known. The exact quasiclassical answer is also know.[5–8] It follows that these problems will serve as a good check on the validity of the variational method as applied to instantons.

We shall see that even very simple trial action S_1 will lead to quite accurate level splitting and band gap. We shall examine two modifications of the variational principle (Secs. 2 and 3) and derive a number of general formulas that can be used for a wide class of potentials in quantum mechanics. Our theory is much simpler than the quasiclassical approach to the functional integral because the most laborious step in this method, namely, the evaluation of the one-instanton determinant, is no longer necessary.

2. TWO-WELL PROBLEM. VARIATIONAL PRINCIPLE I

As a simple illustration of the ideas given in the Introduction, let us consider the well-known problem[5–7] of the determination of the splitting of low-lying levels in the two-well potential:

$$V(x)=\lambda(x^2-\eta^2)^2,\quad \omega_0^2=8\lambda\eta^2. \tag{6}$$

It is well known[5–7] that the level splitting ΔE can be expressed in terms of the functional integral

$$\exp\left\{-T\left(-\frac{\Delta E}{2}\right)\right\}=\int Dx(t)\exp\left\{-\int_0^T dt\left[\frac{\dot{x}^2}{2}+V(x)\right]\right\}$$
$$\times\left\{\int Dx(t)\exp\left[-\int_0^T dt\left(\frac{\dot{x}^2}{2}+\frac{\omega_0^2x^2}{2}\right)\right]\right\}^{-1}, \tag{7}$$

where we have transformed to the Euclidean formulation, i.e., to imaginary time. Moreover, to obtain the lowest-level shift for $T\to\infty$, which is equal to half the level splitting ΔE, we have divided by the "free" partition function of the harmonic oscillator of frequency ω_0 that corresponds to the expansion of the potential (6) in each of the wells.

Let us recall how (7) can be evaluated in the quasiclassical approach. The equation of motion corresponding to the action

$$S=\int dt[\dot{x}^2/2+V(x)], \tag{8}$$

has the solution

$$x_0(t-\tau)=\eta\ \mathrm{th}\ [\omega_0(t-\tau)/2], \tag{9}$$

which is called an instanton, where τ is the center of the instanton. The instanton is a classical subbarrier path in imaginary time that corresponds to a transition from the left-hand well ($x=-\eta$) to the right-hand well ($x=\eta$). The anti-instanton differs from (9) only by the sign of the expression, and corresponds to the reverse transition. The classical action (8) on the instanton (9) is

$$S_0={}^4/_3(2\lambda)^{1/2}\eta^3=\omega_0^3/12\lambda. \tag{10}$$

This is the quasiclassical parameter and is assumed to be large. Expanding (8) around the instanton (9), and neglecting terms higher than the quadratic term in $y=x-x_0$, we obtain

$$S\approx S_0+\int dt[{}^1/_2\dot{y}^2(t)+{}^1/_2V''(x_0(t-\tau))y^2(t)]. \tag{11}$$

The next step is to integrate the numerator in (7) over the fluctuations $y(t)$ with the Gaussian weights (11), and divide by the analogous functions integral for the harmonic oscillator [see (7)]. Since the quadratic form given by (11) clearly has a zero mode connected with the translational invariance of the instanton (9), i.e., with the possibility of taking the instanton center τ at any time, the Gaussian integral over $y(t)$ diverges, which is reflected in the form of the integral with respect to the instanton center τ (the transformation Jacobian between integration over the Fourier coefficient of the zero mode to integration with respect to the instanton center τ is $\omega_0(S_0/2\pi)^{1/2}$). The remaining determinant for the nonzero modes to the power $-1/2$ can be evaluated exactly for this problem (the procedure is quite laborious) and is given by

$$d=\sqrt{12}. \tag{12}$$

Since the instanton-anti-instanton transitions occur many times and at arbitrary instants of time, we obtain the following expression for (7) (the factorial factor appears in the course of transformation from the ordered integration over the centers to integration over the entire region):

$$\exp\left(T\frac{\Delta E}{2}\right)\approx\sum_{N=0}^{\infty}\frac{1}{(2N)!}\int_0^T d\tau_1\ldots d\tau_{2N}$$
$$\times\left\{\left[\omega_0\left(\frac{S_0}{2\pi}\right)^{1/2}\right]de^{-S_0}\right\}^{2N}$$
$$=\mathrm{ch}\left[\omega_0T\left(\frac{S_0}{2\pi}\right)^{1/2}de^{-S_0}\right]\approx\exp\left[T\omega_0\left(\frac{S_0}{2\pi}\right)^{1/2}de^{-S_0}\right],$$

and hence the splitting is given by

$$\Delta E=2\omega_0(S_0/2\pi)^{1/2}e^{-S_0}\sqrt{12},\quad S_0=\omega_0^3/12\lambda. \tag{13}$$

Our problem is to reproduce this result approximately within the framework of the variational approach. Essentially, we have to perform an approximate evaluation of the functional determinant (12). We shall see that our proposed method will also be suitable in other problems for which the exact evaluation of the determinant is not possible.

It is clear from physical considerations that the main contribution to the functional integral (7) is provided by subbarrier transitions from one well to the other, and vice versa. Let $x_0(t-\tau)$ represent the corresponding path with the center at time τ. We shall not assume that $x_0(t-\tau)$ will have the specific instanton form given by (9). We shall only suppose that $x_0(\pm\infty)=\pm\eta$. Thus, we take the trial function $\bar{x}(t)$ in the form

$$\bar{x}(t,\tau_1\ldots\tau_{2N})=\sum_{k=1}^{2N}(-1)^k x_0(t-\tau_k). \tag{14}$$

This formula describes an alternation of instantons and anti-instantons with centers at the points $\tau_1 \ldots \tau_{2N}$. To estimate the contribution of the configuration (14) to (7), we construct the approximate action

$$e^{-S_1[x]}=\frac{1}{2N!}\int_0^T d\tau_1\ldots d\tau_{2N}(De^{-S_0})^{2N}$$

$$\times\exp\{-W[x(t)-\bar{x}(t,\tau_1\ldots\tau_{2N})]\}, \quad (15)$$

where S_0 is the classical action on the trial path x_0, D is the parameter to be varied, and $W[x-\bar{x}]$ is a quadratic form that limits the fluctuations around the trial path $\bar{x}$. In (15), we have neglected fluctuations in the number N of instantons, and consider only one term for given N (which will also be varied). It is well known that this is valid in the "thermodynamic limit" $T\to\infty, N\to\infty, N/T=$ const. We now take the functional W in the simplest form:

$$W[y]=W_0[y]=\int dt\,(1/_2\dot{y}^2+1/_2\omega_0^2y^2). \quad (16)$$

This choice is dictated by the fact that, well away from the instanton centers, W_0 gives the correct description of small oscillations with frequency ω_0 in one of the wells. At the time of the subbarrier transition itself, small oscillations will, of course, have other frequencies, so that (15) gives us only an approximation to the exact action.

The simplest approach is to use (15) to determine the approximate partition functions (divided by the partition function for the harmonic oscillator)

$$\frac{Z_1}{Z_0}=\frac{\int\exp(-S_1[x])Dx}{\int\exp(-W_0[x])Dx}=\frac{1}{2N!}(DTe^{-S_0})^{2N}, \quad (17)$$

where the integral in the numerator is evaluated by changing the variable so that $x=\bar{x}+y$, after which it cancels out with the denominator.

In accordance with the variational procedure, we must now correct the error introduced by using the approximate S_1 and evaluate

$$\langle S-S_1\rangle=\frac{1}{Z_1}\int Dxe^{-S_1[x]}(S[x]-S_1[x])$$

$$=\frac{1}{Z_1}\int d\tau_1\ldots d\tau_{2N}\frac{1}{2N!}(De^{-S_0})^{2N}\int dye^{-W[y]}$$

$$\times\{S[\bar{x}(t,\tau_1\ldots\tau_{2N})+y(t)]-S_1[\bar{x}(t,\tau_1\ldots\tau_{2N})+y(t)]\}, \quad (18)$$

where we have shifted the integration variable $x=\bar{x}+y$. Let us expand $S[\bar{x}+y]$ in powers of y. We note that the zero-order term in the expansion can be written in the form $S[\bar{x}]=2NS_0$ because the gas is tenuous (this is controlled by the parameter $\omega_0^3/12\lambda$), the linear term in y is annulled as a result of integration with respect to y with the even weight $\exp[-W_0(y)]$ (independently of whether or not $\bar{x}$ satisfies the equation of motion), and the quadratic term in y is

$$\frac{1}{2}\int dt[V''(\bar{x}(t))y^2(t)+\dot{y}^2(t)].$$

This expression is averaged in (18) over the centers of the tenuous instanton gas. Since the particle is mostly found in one of the wells, where the frequency is ω_0, and in the tenuous gas all the centers τ_i are, on average, at a large distance from one another, we have the approximate result

$$\frac{1}{2}\int dt[V''(\bar{x}'(t,\tau_1\ldots\tau_{2N}))y^2+\dot{y}^2]$$

$$\approx\frac{1}{2}\int dt[(\omega_0^2-2N\omega_1^2)y^2+\dot{y}^2], \quad (19)$$

where

$$\omega_1^2=\omega_0\int_{-\infty}^{+\infty}dt[\omega_0^2-V''(x_0(t-\tau))]. \quad (20)$$

Let us now evaluate the second term in braces in (18). According to the definition given by (15)

$$\exp\{-S_1[\bar{x}(t,\tau_1\ldots\tau_{2N})+y(t)]\}$$

$$=\frac{1}{2N!}(De^{-S_0})^{2N}\int d\tau_1'\ldots d\tau_{2N}'$$

$$\times\exp\{-W_0[\bar{x}(t,\tau_1\ldots\tau_{2N})+y(t)-\bar{x}(t,\tau_1'\ldots\tau_{2N}')]\}.$$

We shall see below that the main contribution to the integrals with respect to τ_i' is provided by regions where the centers of the primed (anti)instantons are close to the centers of the unprimed (anti)instantons. If we consider the different ways of comparing primed and unprimed centers, we find that there are $2N!$ such centers. Expanding the argument of the functional W_0 in powers of $\tau_i'-\tau_i=\delta_i$ we obtain

$$W_0[\bar{x}+y-\bar{x}']=\frac{1}{2}\int dt\left[y(t)+\sum_{i=1}^{2N}(-1)^i\dot{x}_0(t-\tau_i)\right]$$

$$\times\left(\omega_0^2-\frac{\partial^2}{\partial t^2}\right)\left[y(t)+\sum_{j=1}^{2N}(-1)^j\dot{x}_0(t-\tau_j)\delta_j\right]. \quad (21)$$

We thus have a Gaussian integral over the displacements δ_i. We now substitute

$$\gamma=\int dt[\ddot{x}_0^2(t)+\omega_0^2\dot{x}^2(t)]. \quad (22)$$

This quantity defines the "sharpness" of the integrals over δ_i and is large:

$$\gamma\sim\omega_0^2(\omega_0^3/\lambda)\gg\omega_0^2.$$

This enables us to confine (21) to the terms written out in that expression.

Neglecting overlap integrals corresponding to instantons at different centers, we obtain the following expression by shifting the Gaussian integrals over δ_i:

$$S_1[\bar{x}(t,\tau\ldots)+y(t)]=-\ln\left[(De^{-S_0})^{2N}\left(\frac{2\pi}{\gamma}\right)^N\right]$$

$$+\frac{1}{2}\int dty\left(\omega_0^2-\frac{\partial^2}{\partial t^2}\right)y$$

$$+\frac{1}{2\gamma}\sum_{i=1}^{2N}\left[\int dty\left(\omega_0^2-\frac{\partial^2}{\partial t^2}\right)\dot{x}_0(t-\tau_i)\right]^2.$$

The last term in this expression must also be averaged over the statistical ensemble, i.e., we must integrate over all the τ_i. We thus see that the evaluation of $\langle S-S_1\rangle$ now reduces to integration of a quadratic expression over the fluctuations $y(t)$ with a weight equal to an exponential whose argument is equal to the harmonic-oscillator action. This integration can be performed in an elementary manner, since

$$\langle y(t_1)y(t_2)\rangle=\frac{\int Dy\exp(-W_0[y])y(t_1)y(t_2)}{\int Dy\exp(-W_0[y])}=\mathscr{D}(t_2-t_1)$$

is the harmonic-oscillator propagator:

$$\mathscr{D}(t)=\int\frac{e^{i\omega t}}{\omega^2+\omega_0^2}\frac{d\omega}{2\pi},\quad \mathscr{D}(0)=\frac{1}{2\omega_0}. \tag{23}$$

Using this, we obtain

$$\langle S-S_1\rangle=2N\left[\ln D\left(\frac{2\pi}{\gamma}\right)^{1/2}-\frac{\omega_1^2}{4\omega_0^2}+\frac{1}{2}\right].$$

Recalling the expression for Z_1/Z_0, given by (17), we obtain

$$\frac{Z}{Z_0}\geqslant\frac{Z_1}{Z_0}e^{-\langle S-S_1\rangle}=\exp\left\{-2N\ln 2N+2N+2N\left[\ln\left(Te^{-S_0}\left(\frac{\gamma}{2\pi}\right)^{1/2}+\frac{\omega_1^2}{4\omega_0^2}-\frac{1}{2}\right]\right\}, \tag{24}$$

where we note that the parameter D introduced above has canceled out. Taking the maximum of this expression in N, we obtain finally

$$\Delta E\geqslant 2\left(\frac{\gamma}{2\pi}\right)^{1/2}\exp\left(\frac{\omega_1^2}{4\omega_0^2}-\frac{1}{2}\right)e^{-S_0}, \tag{25}$$

where the quantities S_0, ω_1^2 and γ, all of which are functionals of the trial function $x_0(t)$, are given by (8), (20), and (22), respectively.

We must now find the maximum of the right-hand side of (25) by performing a variation over the trial subbarrier path $x_0(t)$. It is clear, by the way, that, in the quasiclassical limit, the maximum is determined by the minimum of the action S_0. We emphasize that the above formula is valid for a two-well problem of any profile.

It is useful to compare (25) with the exact quasiclassical formula (13). The factor $(\gamma/2\pi)^{1/2}$ can be interpreted as the determinant of the transformation from integration over the zero mode to integration over τ. In (13), this applies to $(S_0/2\pi)^{1/2}$. The exponential factor in (25), which corresponds to the determinant for the nonzero modes, represents the change in the frequency of small oscillations at the time of the instanton transition: according to (20), ω_1 is, precisely, the change in frequency averaged over the transition time.

Specifically for the potential (6), the maximum is reached on the instanton (9). Evaluation of the simple integrals in (8), (20), and (22) yields the following expressions for this case:

$$S_0=\omega_0^3/12\lambda,\quad \omega_1^2=6\omega_0^2,\quad \gamma=\omega_0^5/10\lambda.$$

Substituting these values in (24), we obtain

$$\Delta E\geqslant 2\omega_0\left(\frac{\omega_0^3}{2\pi\lambda}\right)^{1/2}\exp\left(-\frac{\omega_0^3}{12\lambda}\right)\left(\frac{e^2}{10}\right)^{1/2}. \tag{26}$$

The last factor, $(e^2/10)^{1/2}\approx 0.86$), presents us with a difference as compared with the exact formula (13). Thus, (26) correctly reproduces all the dependences on the parameters of the problem, and the numerical coefficient in this expression differs by only 14% from the exact answer. For a periodic potential, which we shall examine by a somewhat different method in the next section (and which is closer to real problems in field theory), the error is smaller still, i.e., only about 5%. The difference in accuracy is probably due to the presence in the exact quadratic form of the two-level problem of an exta discrete level whose contribution is not satisfactorily taken into account by the action S_1. We draw attention to the fact that we did not have to solve the most laborious part of the problem, namely, the evaluation of the one-instanton determinant. Once the general formula given by (15) has been derived, all that remains is to evaluate a few simple integrals.

3. PERIODIC POTENTIAL. VARIATIONAL PRINCIPLE II

We now consider the quantum mechanics of a particles in a periodic potential of the form

$$V(x)=(\omega_0^2/2\lambda^2)\cos^2\lambda x \tag{27}$$

in the quasiclassical limit $\omega_0/\lambda^2\gg 1$, in which the width of the lowest allowed band is exponentially small and can be calculated by using the instanton-anti-instanton gas that describes tunneling between the wells.[8] The basic problem here is again the evaluation of the one-insanton determinant. We note that the solutions of the Schrödinger equation with the potential given by (27) are Mathieu functions, and the band widths can be found in Ref. 9. The energy of a state in the lowest band with quasimomentum θ $(-\pi<\theta<\pi)$ is given by

$$E(\theta)=\omega_0/2-\cos\theta 4\omega_0(\omega_0/\pi\lambda^2)^{1/2}\exp(-2\omega_0/\lambda^2). \tag{28}$$

On the other hand, this result must also be obtained for $T\to\infty$ from the path integral:

$$\exp\left\{-T\left[E(\theta)-\frac{\omega_0}{2}\right]\right\}=\int Dx\exp\left\{-\int_0^T dt\left[\frac{\dot x^2}{2}+V(x)\right]+i\theta\int_0^T\dot x\,dt\right\}\left\{\int Dx\exp[-W_0(x)]\right\}^{-1}, \tag{29}$$

where the quasimomentum is introduced by means of the so-called θ-term.

It is clear from physical considerations that the main contribution to the functional integral in (28) should be provided by the path in the form of an instanton-antiinstanton gas:

$$\bar x(t,\tau_1\ldots\tau_{N_+},\bar\tau_1\ldots\bar\tau_{N_-})=\sum_{i=1}^{N_+}x_0(t-\tau_i)-\sum_{j=1}^{N_-}x_0(t-\bar\tau_j), \tag{30}$$

where $x_0(\pm\infty)=\pm\pi/2\lambda$, but, otherwise, $x_0(t)$ is so far arbitrary.

We must now choose the approximate action. We have seen that the imprecision of variational principole I, examined in the last section, was connected with the fact that the quadratic form $W_0[y]$ does not correctly represent small oscillations around the path at the time of the transition. It is therefore better to use instead the exact quadratic form of the instanton action (11):

$$W[y]=\frac{1}{2}\int dt[\dot{y}^2+V''(\bar{x}(t))y^2]. \tag{31}$$

However, because of the absence of zero modes, this action does not limit the deviation corresponding to the shift of the instanton centers on the path (30). This difficulty can be overcome with the aid of a device analagous to the Faddeev-Popov method in gauge theory (see, for example, Ref. 5). Let us subject the deviation $y=x-\bar{x}$ to the zero-mode orthogonality condition by inserting the following "unity" into the path integral:

$$\int d\tau_1\ldots d\tau_{N_+}d\bar{\tau}_1\ldots d\bar{\tau}_{N_-}\delta\left[\int\psi(t-\tau_i)y(t)\,dt\right]\times\left[\int dt\psi(t)\dot{x}_0(t)\right]^{N_++N_-}, \tag{32}$$

where $\psi(t)$ is any function that is not orthogonal to the zero mode $\dot{x}_0$. In practice, it is more convenient to smear out the δ-function with a Gaussian weight and, instead of (32), introduce

$$\int\prod_{i=1}^{N_++N_-}d\tau_i\left[\frac{\eta}{\sqrt{2\pi}}\int dt\psi\dot{x}_0\right]^{N_++N_-}\times\exp\left\{-\frac{\eta^2}{2}\sum_i\left[\int dt\psi(t-\tau_i)y\right]^2\right\}. \tag{33}$$

After integration over y, this expression is found to be independent of both the parameter η and the specific choice of ψ.

We therefore take the approximate action in the form

$$\exp(-S_1[\bar{x}])=\frac{e^{i\theta(N_+-N_-)}}{N_+!N_-!}\int d\tau_1\ldots d\tau_{N_+}d\bar{\tau}_1\ldots d\bar{\tau}_{N_-}\left(\frac{\eta d}{\sqrt{2\pi}}\right)^{N_++N_-}$$
$$\times\exp\left\{-S[\bar{x}]-W[x-\bar{x}(t,\tau\ldots)]-\frac{\eta^2}{2}\sum_i\left[\int dt\psi(t-\tau_i)(x-\bar{x}(t,\tau\ldots)\right]^2\right\}, \tag{34}$$

where $W[y]$ is given by (31). This expression now restricts any small deviation from the trial path $\bar{x}(t,\tau,\ldots)$, and includes the variational parameter d.

The simplest thing to do is to use S_1 to determine the average $\langle S-S_1\rangle$:

$$\langle S-S_1\rangle=\frac{1}{Z_1}\int\prod d\tau\,d\bar{\tau}'\left(\frac{\eta d}{\sqrt{2\pi}}\right)^{N_++N_-}\int dy\exp\left\{-S[\bar{x}]-W[y]-\frac{\eta^2}{2}\sum_i\left[\int dt\psi(t-\tau_i)y(t)\right]^2\right\}$$
$$\times\{S[\bar{x}(t,\tau\ldots)+y]-S_1[\bar{x}(t,\tau\ldots)+y]\}. \tag{35}$$

Let us expand the exact action in terms of the small quantity y and recall that the linear term need not be written out because it is annulled on integration over y:

$$S[\bar{x}+y]\approx S[\bar{x}]+W[y]-i\theta(N_+-N_-). \tag{36}$$

Next, consider the second term in the last pair of braces in (35). According to (34), we have

$$\exp\{-S_1[\bar{x}(t,\tau\ldots)+y]\}=e^{i\theta(N_+-N_-)}\left(\frac{\eta d}{\sqrt{2\pi}}\right)^{N_++N_-}$$
$$\times\int\Pi\,d\tau'\,d\bar{\tau}'\exp\left\{-S[\bar{x}]-W\left[\sum_i\dot{x}_0(t-\tau_i)(\tau_i-\tau_i')+y\right]-\frac{\eta^2}{2}\sum_i\left[\int dt\psi(t-\tau_i)\left(\sum_j\dot{x}_0(t-\tau_j)(\tau_j-\tau_j')+y\right)\right]^2\right\}. \tag{37}$$

As can be seen, we have expanded the difference $\bar{x}(t,\tau'\ldots)-\bar{x}(t,\tau\ldots)$ and have canceled out $N_+!N_-!$ because this is the number of ways of comparing the primed and unprimed sets of centers.

We note that $W[y+\ldots]$ does not contain the dependence on τ' because $\dot{x}_0(t-\tau_i)$ are zero modes for this quadratic form. Integration over τ' can be performed in an elementary manner if we neglect overlap integrals between different instantons. We thus obtain

$$\exp(-S_1)=\exp\{i\theta(N_+-N_-)\}\left(\frac{d}{\int dt\psi(t)\dot{x}_0(t)}\right)^{N_++N_-}\times\exp\{-S[\bar{x}]-W[y]\}, \tag{38}$$

and, if we compare this with (36), we see that, by taking [in accordance with (33)!]

$$d=\int\psi(t)\dot{x}_0(t)\,dt, \tag{39}$$

we ensure that the difference $S-S_1$ vanishes before integration over y and τ_i. It is, of course, possible to keep d as the parameter to be varied and eventually find the maximum with respect to d. However, the final answer will be the same.

We see that, in this modification of the variational principle, we have selected a very good approximation for the action ($\langle S-S_1\rangle=0$). We now have to pay the price for this by having to perform a relatively complicated evaluation of Z_1 (we recall that the reverse situation was encountered in the case of variational principle I).

Thus, we must examine the approximate partition function. In accordance with the definition give by (34), we have

$$Z_1=\frac{e^{i\theta(N_+-N_-)}}{N_+!N_-!}\left(\frac{\eta Td}{\sqrt{2\pi}}e^{-S[x_0]}\right)^{N_++N_-}\int Dye^{-F[y]}, \tag{40}$$

where $F[y]$ is a functional of the effective action, averaged over the statistical ensemble:

$$e^{-F[y]}=\frac{1}{T^{N_++N_-}}\int\prod d\tau\,d\bar{\tau}\exp\left\{-W[y]-\frac{\eta^2}{2}\sum_i\left(\int dt\psi(t-\tau_i)y\right)^2\right\}. \tag{41}$$

If we were to perform the statistical averaging exactly, and then evaluate exactly the integral over y, we would obtain the exact quasiclassical answer. Since this is not our problem, we exploit once again the convexity property of the exponential, and take the average of the action in (41) in the exponential.

We obtain

$$\exp(-F[y])\geqslant\exp\Big\{-\langle W[y]\rangle - \frac{\eta^2}{2}\sum_i\Big\langle\Big(\int dt\psi(t-\tau_i)y(t)\Big)^2\Big\rangle\Big\}, \tag{42}$$

where the angle brackets represent averaging over the statistical ensemble. We shall perform this averaging by using the fact that the instanton gas is tenuous:

$$\langle W[y]\rangle=\frac{1}{2}\int dt(\dot y^2+\omega_0^2y^2) + \frac{1}{2}\int dty^2\langle[V''(\bar x(t,\tau\dots))-\omega_0^2]\rangle$$

$$\approx W_0[y]+\frac{1}{2}\frac{N_++N_-}{T}\int d\tau[V''(x_0(t-\tau))-\omega_0^2]\int dty^2(t)$$

$$=W_0[y]-\frac{1}{2}\frac{N_++N_-}{T}\frac{\omega_1^2}{\omega_0}\int dty^2(t), \tag{43}$$

where, in the last equation, we have used (20). We thus see that the average change in the frequency of oscillations corresponding to the instanton transition ω_1 has appeared once again. The second term in (42) can be averaged in a similar way, and we obtain the effective action $F[y]$ in this approximation:

$$F[y]\approx W_0[y]-\frac{N_++N_-}{T}\frac{\omega_1^2}{\omega_0}\int dty^2(t) + \frac{\eta^2(N_++N_-)}{T}\int dt_1\,dt_2y(t_1)y(t_2)G(t_1-t_2), \tag{44}$$

$$G(t_1-t_2)=\int d\tau\psi(t_1-\tau)\psi(t_2-\tau), \tag{45}$$

i.e., in this approximation, the effective action is quadratic in the deviation y. In the general case, it will contain all powers of y.

The Gaussian integral (40) with the action given by (44) can be evaluated in an elementary manner if we transform to Fourier components. The result is

$$\frac{Z_1}{Z_0}\geqslant\frac{e^{i\theta(N_+-N_-)}}{N_+!N_-!}\Big[\Big(\frac{\eta Td}{\sqrt{2\pi}}\Big)e^{-S[x_0]}\Big]^{N_++N_-}\exp\Big\{-\frac{T}{2}\int_{-\infty}^{+\infty}\frac{d\omega}{2\pi} \times\ln\Big[1+\frac{N_++N_-}{T}\frac{\eta^2G(\omega)-\omega_1^2/\omega_0^2}{\omega^2+\omega_0^2}\Big]\Big\}. \tag{46}$$

Since N/T is a small parameter for the tenuous gas, we can expand the last factor. Recalling (39), we have

$$\frac{Z_1}{Z_0}\geqslant\frac{e^{i\theta(N_+-N_-)}}{N_+!N_-!}\Big[\frac{\eta T}{\sqrt{2\pi}}\Big(\int\frac{d\omega}{2\pi}\psi(\omega)\dot x_0(-\omega)\Big)e^{-S_0} \times\exp\Big\{\frac{\omega_1^2}{4\omega_0^2}-\frac{\eta^2}{2}\int\frac{d\omega}{2\pi}\frac{G(\omega)}{\omega_0^2+\omega^2}\Big\}\Big]^{N_++N_-} \tag{47}$$

By varying this expression with respect to η and $\psi(\omega)$, we obtain the best value:

$$\eta=\Big(\int\frac{d\omega}{2\pi}\frac{G(\omega)}{\omega^2+\omega_0^2}\Big)^{-1/2},\quad \psi(\omega)=(\omega^2+\omega_0^2)\dot x_0(\omega). \tag{48}$$

It is clear that the best $\psi(t)$ is not at all identical with the zero mode $\dot x_0(t)$. The difference between these functions effectively corrects the nonzero-mode determinant. Substituting (48) in (47), we obtain

$$\frac{Z_1}{Z_0}=\frac{e^{i\theta(N_+N_-)}}{N_+!N_-!}\Big[T\Big(\int\frac{dt}{2\pi}(\ddot x^2+\omega_0^2\dot x^2)\Big)^{1/2} \times\exp\Big(\frac{\omega_1^2}{4\omega_0^2}-\frac{1}{2}\Big)e^{-S_0}\Big]^{N_+N_-}. \tag{49}$$

We note that this is literally identical with (24), which was deduced from variational principle I (with the exception of factorial factors, whose presence is due to the somewhat different formulation of the problem for the periodic potential). To obtain the energy as a function of the quasimomentum θ, we can either sum (49) over $N_\pm$, or find and substitute their extremal values. In either case, we obtain

$$E(\theta)=\frac{\omega_0}{2}-\cos\theta 2\Big(\frac{\gamma}{2\pi}\Big)^{1/2}\exp\Big(\frac{\omega_1^2}{4\omega_0^2}-\frac{1}{2}\Big)e^{-S_0}, \tag{50}$$

where S_0 is the action for the classical subbarrier path, and the functionals ω_1 and γ are, as before, given by (20) and (22). Specifically for the potential (27), we have an instanton of the form

$$x_0(t-\tau)=\frac{2}{\lambda}\operatorname{arctg}\Big(e^{\omega_0(t-\tau)}-\frac{\pi}{4}\Big),$$

$$\dot x_0(t-\tau)=\frac{\omega_0}{\lambda}\frac{1}{\operatorname{ch}\omega_0(t-\tau)}, \tag{51}$$

and hence

$$S_0=2\omega_0/\lambda^2,\quad \omega_1^2=4\omega_0^2,\quad \gamma=8\omega_0^3/3\lambda^2.$$

Substituting these values in (50), and comparing with the exact formula (28), we see that our approximate answer differs from the exact answer by the factor $(e/3)^{1/2}\approx0.95$, i.e., the error in the approximate calculation of the determinant is 5%.

We now see that, in a certain sense, the variational principle I examined in Sec. 2 is the first approximation to variational principle II of the present section. It is obtained if we use (42) in the course of averaging over the statistical ensemble. Equation (42) can be improved by using a formula such as (5) in the averaging process. However, we then lose the above result for the lower bound of the partition function. We are therefore not entitled to vary again with respect to η and $\psi(t)$, and must use the values obtained from the first approximation. On the other hand, since they were chosen so as to obtain the best approximation to $F[y]$, the numerical series in powers of $F-F^{(1)}$ should converge quite rapidly.

We shall not reproduce the details of the calculation, and quote only the final formula for the second approximation. It is simply related to the level splitting (or the splitting of a wide band) in the first approximation:

$$\Delta E^{(2)}=\Delta E^{(1)}e^{C-1/4},\quad C=\frac{1}{4}\int\frac{d\omega}{2\pi}\frac{f^2(\omega)}{\omega_0(\omega^2+4\omega_0^2)}$$

$$f(\omega)=\int dte^{-i\omega t}[\omega_0^2-V''(x_0(t))].$$

The last quantity is the Fourier transform of the change in the frequency of small oscillations around the instanton path. The values of ΔE, obtained in the first and second approximation, are listed in the following table as fractions of the exact values:

	Two wells	Periodic potential
First approximation	0.86	0.95
Second approximation	0.96	0.991

The basic aim of the present paper is methodological, i.e., to learn how to evaluate the contribution to the functional integral due to the neighborhood of instanton-type paths that do not necessarily satisfy the equations of motion. We have used the above quantum-mechanical problems to verify that the variational principle yields good accuracy as compared with observed values. The generalization of this method to the case of quantum-field theory will be considered in another paper.

We are indebted to L. N. Lipatov and A. V. Turbiner and, especially, to E. L. Kremer and M. I. Éĭdes for numerous fruitful discussions.

[1)]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. M. Tyupkin, Phys. Lett. **59**, 85 (1975).

[2)]G. S. Callan, R. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978); *ibid.* **19**, 1826 (1979).

[3]A. P. Bukhvostov and L. N. Lipatov, ICTP (Trieste), preprint 80/2, 1980.

[4]R. Feynman and A. R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill, 1965 [Russian translation, Mir, Moscow, 1968].

[5]A. M. Polyakov, Nucl. Phys. B **120**, 429 (1977).

[6]S. Coleman, The Uses of Instantons, Preprint HUTP-788/A004.

[7]A. I. Vaĭnshteĭn, V. I. Zakharov, A. V. Novikov, and M. A. Shifman, Usp. Fiz. Nauk **136**, 553 (1982) [Sov. Phys. Usp. **25**, 195 (1982)].

[8]M. Stone, Phys. Rev. D **18**, 4752 (1978).

[9]Handbook of Mathematical Functions, ed. by M. Abramovitz and I. E. Stegun, Dover, 1964 [Russian translation, Nauka, Moscow, 1979, p. 537].

Translated by S. Chomet