Properties of cosmologies with dynamical pseudo Nambu-Goldstone bosons

S. C. Cindy Ng* and David L. Wiltshire†
Department of Physics and Mathematical Physics, University of Adelaide, Adelaide, S.A. 5005, Australia
(Received 12 April 2000; published 18 December 2000)

We study observational constraints on cosmological models with a quintessence field in the form of a dynamical pseudo Nambu-Goldstone boson. After reviewing the properties of the solutions, from a dynamical systems phase space analysis, we consider the constraints on parameter values imposed by luminosity distances from the 60 type Ia supernovae published by Perlmutter *et al.*, and also from gravitational lensing statistics of distant quasars. In the case of the type Ia supernovae we explicitly allow for the possibility of evolution of the peak luminosities of the supernovae sources, using simple empirical models which have been recently discussed in the literature. We find weak evidence to suggest that the models with supernovae evolution fit the data better in the context of the quintessence models in question. If source evolution is a reality then the greatest challenge facing these models is the tension between the current value of the expansion age, $H_0 t_0$, and the fraction of the critical energy density, $\Omega_{\phi 0}$, corresponding to the scalar field. Nonetheless there are ranges of the free parameters which fit all available cosmological data.

DOI: 10.1103/PhysRevD.63.023503 PACS number(s): 98.80.Cq, 95.35.+d, 98.80.Es

I. INTRODUCTION

Scalar fields have played a central role in models of the very early universe for the past 20 years. In the past few years attention has turned to models in which a scalar field plays a dynamical role at late times, rather than simply being frozen in as a static relic vacuum energy. Such models, which have been dubbed ''quintessence'' models [1], could in principle provide a dynamical solution to the cosmological constant problem—namely the question of why the magnitude of the vacuum energy at the present epoch is so much smaller than one might naively expect from particle physics models such as various supergravity theories. A dynamical ''solution'' of the cosmological constant problem would amount to a demonstration that a particular dynamical evolution of the scalar quintessence field is a natural consequence of the cosmological field equations without fine-tuning of parameters, given some reasonable physical assumptions about the initial conditions.

The most notable recent observational evidence which has driven the theoretical interest is the measurement of the apparent magnitude-redshift relationship using type Ia supernovae (SNe Ia) [2]. These results have been interpreted, in the context of a cosmological model containing pressureless dust and a cosmological constant Λ, as evidence that the universe is undergoing accelerated expansion at the present epoch (see [3,4] and references therein). The validity of this conclusion is currently open to some doubt, however. In particular, a recent analysis by Riess *et al.* [5] indicates that the sample of type Ia supernovae shows a possible evolution in rise times from moderate ($z \sim 0.3$) to large ($z \sim 1$) redshifts. Although the statistical significance of this result has been diminished—from the 5.8σ level [5] to the 1.5σ level [6]—upon a more rigorous treatment of the uncertainties in the data [6], it remains true that while a systematic evolution in the rise times of the supernovae is not conclusively ruled in, neither is it conclusively ruled out.

Given that an evolution in the shape of the light curves of the supernovae measured in their rest frame remains a real possibility, it would not be surprising if the peak luminosity—which is the effective standard candle used—were also to evolve. Riess *et al.* [5] conclude that the type Ia supernovae data could conceivably be explained entirely within the context of an open Friedmann-Robertson-Walker universe together with a reasonable astrophysical evolution model, e.g., a consequence of a time variation of the abundances of relevant heavy elements in the environment of the white dwarf supernovae progenitors. Detailed astrophysical modeling—see, e.g., [7]—should hopefully eventually resolve the issue, although at this stage the difference between our theoretical understanding and the observations remains quite substantial [8].

In many recent papers it has been commonly assumed that the dynamical scalar field, ϕ, should obey an effective equation of state $P_\phi \simeq w \rho_\phi$ with $-1 < w < 0$, at the present epoch, in order to obtain a cosmological acceleration, i.e., a negative deceleration parameter q_0. Indeed, the condition that $-1 < w < 0$ is often taken as a defining characteristic of ''quintessence'' [1]. The broad picture in this cosmological scenario is that the universe is currently in the early stage of an epoch of inflationary expansion. The motivation for this is that one could then hope to have a model cosmology in which observations such as the type Ia supernovae apparent magnitude-redshift relation could be explained by a cosmological acceleration in a similar fashion to models with a cosmological constant, but with the possibility of explaining why the magnitude of the vacuum energy density and the energy density in ordinary pressureless matter, ρ_m, are comparable at the present epoch—the so-called ''cosmic coincidence problem'' [9].

One attractive feature of homogeneous isotropic cosmological models with dynamical scalar fields is that many of them possess ''cosmological scaling solutions'' [10], namely solutions which at late times have energy density compo-

*Electronic address: cng@physics.adelaide.edu.au
†Electronic address: dlw@physics.adelaide.edu.au

nents which depend on the cosmic scale factor according to $\rho \propto a^{-m_1}$ and $\rho_\phi \propto a^{-m_2}$ simultaneously, and which act as attractors in the phase space. If $m_2 < m_1$, which is the case, for example, for simple power-law potentials with inverse powers [11,12] $V(\phi) \propto \phi^{-\alpha}$, or for certain power-law potentials with positive powers [10], then the scalar field dominates at late times, producing a quintessence-dominated cosmology with accelerated expansion at late times. If $m_1 = m_2$, which is the case for exponential potentials [13–20], then the scaling solutions are ''self-tuning'' [16]—i.e., the dynamics of the scalar field follows that of the other dominant energy component, with a dependence $\rho_\phi \propto a^{-4}$ in the radiation-dominated era and a dependence $\rho_\phi \propto a^{-3}$ in the matter-dominated era.

Even if the ultimate late-time properties of the solutions are not precisely ''self-tuning'' in the above sense, models such as those with inverse power law potentials can still effectively act as ''tracking solutions,'' since for a wide range of initial conditions, the solutions rapidly converge to a common, cosmic evolutionary track [12]. Thus there are a number of ways in which one might hope to solve the ''cosmic coincidence problem,'' though in practice a degree of tuning of the parameters has been necessary in all models studied to date.

In this paper, we consider a form of quintessence, an ultra-light pseudo Nambu-Goldstone boson (PNGB) [21] which is still relaxing to its vacuum state. From the viewpoint of quantum field theory PNGB models are the simplest way to have naturally ultra-low mass, spin-0 particles and hence perhaps the most natural candidate for a presently existing minimally coupled scalar field. The effective potential of a PNGB field ϕ can be taken to be of the form [22]

$$V(\phi) = M^4[\cos(\phi/f) + 1], \tag{1}$$

where the constant term is to ensure that the vacuum energy vanishes at the minimum of the potential. This potential is characterized by two mass scales, a purely spontaneous symmetry breaking scale f and an explicit symmetry breaking scale M.

The effective PNGB mass is $m_\phi \sim M^2/f$. To obtain solutions with $\Omega_\phi \sim 1$, the energy scales are essentially fixed [21] to values $M \sim 10^{-3}$ eV, interestingly close to the neutrino mass scale for the Mikheyev-Smirnov-Wolfenstein (MSW) solution to the solar neutrino problem, and $f \sim m_{Pl} \simeq 10^{19}$ GeV, the Planck scale. Since these two energy scales have values which are reasonable from the viewpoint of particle physics, one might hope to explain the coincidence that the vacuum energy is dynamically important at the present epoch.

The cosmology of PNGB models has already been extensively studied in the literature [17,18,20,22–24]. In particular, a number of constraints have been placed on the parameters M and f by various sets of observational data [17,18,23,24]. Most recently, Frieman and Waga [24] have set bounds based on the SNe Ia data of Riess *et al.* [4] (hereafter R98) on the one hand, and gravitational lensing surveys on the other. Comparing these bounds is of interest, since the Sn Ia data have been interpreted as favoring a cosmological constant, whereas gravitational lensing data has been used to place upper bounds on Λ [25–27]. They therefore provide complementary tests of the parameter spaces of models with a non-trivial vacuum energy.

In this paper it is our intention to critically study these bounds. First, we will consider how the bounds are affected by the initial value of the scalar field at the beginning of the matter-dominated epoch. Secondly, we wish to investigate how such bounds might be affected in the case of the PNGB model if the observed apparent faintness of type Ia supernovae is at least partly due to an intrinsic evolution of the sources over cosmological time scales, which in view of the results of [5] would appear to be a very real possibility. The reason for focusing on the PNGB models in such an investigation is suggested by the fact that whereas many quintessence models have been singled out in the literature, perhaps somewhat artificially, simply because they have the property of yielding an accelerated expansion, many different possibilities arise in the PNGB case. Indeed, at very late times, the apparent magnitude-redshift relation for PNGB models ultimately coincides with that of the Einstein–de Sitter model, even though the density of ordinary matter can be low in the PNGB cosmologies. The requirement that the faintness of type Ia supernovae is entirely due to their cosmological distances places rather strong restrictions on the values of the parameters M and f [24], because it requires us to exist at an epoch of the PNGB cosmologies which is still quite far removed from our ultimate destiny. If these restrictions are relaxed because of evolutionary effects, then it is quite plausible that other regions of the parameter space of the PNGB models become viable alternatives. Since PNGB cosmologies could therefore still solve the ''missing energy problem,'' even if the evidence for a cosmological acceleration proves to be ephemeral, we believe it is important to investigate this possibility quantitatively.

We will begin the paper with a qualitative analysis of the solutions, to provide some general insights which will help to guide our quantitative discussion. Although these properties are no doubt already known, to the best of our knowledge an analysis of the phase space of the solutions has never been presented in the literature. Having completed this analysis in Sec. II we will go on to discuss a number of issues relating to numerical integration in Sec. III, and relate the properties of the solutions found numerically to the exact analysis of Sec. II. In Sec. IV we present the main analysis of the constraints imposed on the (M,f) parameter space, allowing for the possibility of evolution of peak luminosities in the type Ia supernova sources. Bounds from gravitational lensing statistics are updated in Sec. V, and the implications of our results are discussed at greater length in Sec. VI.

II. PHASE-SPACE ANALYSIS

We will begin by performing an analysis of the differential equations governing the cosmological evolution in a manner similar to previous studies in inflationary and quintessential models [10,14,15,19,20].

The classical action for gravity coupled to a scalar field ϕ has the form

$$S=\int d^4x\sqrt{-g}\left[\left(\frac{R}{2\kappa^2}-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi-V(\phi)\right)+\mathcal{L}\right], \tag{2}$$

where κ is the Planck constant, R is the Ricci scalar, $g\equiv\det g_{\mu\nu}$, and L is the Lagrangian density of non-relativistic matter and radiation. For simplicity, we assume ϕ is minimally coupled to the curvature, and we work in units in which $\hbar=c=1$.

Consider a spatially flat Friedmann-Robertson-Walker (FRW) universe containing a fluid with barotropic equation of state $P_\gamma=(\gamma-1)\rho_\gamma$, where γ is a constant, $0\leqslant\gamma\leqslant2$, such as radiation ($\gamma=4/3$) or dust ($\gamma=1$). There is a self-interacting scalar field with the PNGB potential energy density (1) evolving in this universe. The total energy density of this homogeneous scalar field is $\rho_\phi=\dot\phi^2/2+V(\phi)$. The governing equations are given by

$$\dot H=-\frac{\kappa^2}{2}(\rho_\gamma+P_\gamma+\dot\phi^2), \tag{3}$$

$$\dot\rho_\gamma=-3H(\rho_\gamma+P_\gamma), \tag{4}$$

$$\ddot\phi=-3H\dot\phi-\frac{dV}{d\phi}, \tag{5}$$

subject to the Friedmann constraint

$$H^2=\frac{\kappa^2}{3}\left(\rho_\gamma+\frac{1}{2}\dot\phi^2+V\right), \tag{6}$$

where $\kappa^2\equiv8\pi G$, $H=\dot a/a$ is the Hubble parameter, and an overdot denotes ordinary differentiation with respect to time t.

We may rewrite the Friedmann constraint as

$$\Omega_\gamma+\Omega_\phi=1 \tag{7}$$

where

$$\Omega_\gamma=\frac{\kappa^2\rho_\gamma}{3H^2}, \tag{8}$$

$$\Omega_\phi=\frac{\kappa^2}{3H^2}\left(\frac{1}{2}\dot\phi^2+V\right) \tag{9}$$

are the ratios of the energy densities of the barotropic matter and the quintessence field as fraction of the critical density respectively.

In contrast to the case of the cosmologies with an exponential potential [14,15,19] where the dynamics can be reduced to a 2-dimensional autonomous phase plane, for the system (3)–(6) the simplest phase space appears to be 3-dimensional in the full four-dimensional phase space.

There are two alternative choices of variables which are useful to describe the dynamics, which we will discuss in turn.

A. Field variables

The first choice is to simply use the Hubble parameter, H, the scalar field and its first derivative as the elementary variables. These are of course simply the variable u,v,w of Frieman and Waga [18] up to an overall scaling. By defining

$$I\equiv\frac{\phi}{f}=\frac{\kappa\phi}{\mathcal{F}};\quad J\equiv\kappa\dot\phi; \tag{10}$$

we therefore obtain the system

$$\dot H=-\frac{3\gamma}{2}H^2+\frac{\gamma m^2}{2}(\cos I+1)-\frac{(2-\gamma)}{4}J^2, \tag{11}$$

$$\dot I=\frac{J}{\mathcal{F}}, \tag{12}$$

$$\dot J=-3HJ+\frac{m^2}{\mathcal{F}}\sin I, \tag{13}$$

where for notational simplicity we define $m^2=\kappa^2M^4$, and $\mathcal{F}=\kappa f$, so that $\mathcal{F}$ is dimensionless, while m has dimensions inverse time. The constraint equation becomes

$$\kappa^2\rho_\gamma=3H^2-m^2(\cos I+1)-\frac{1}{2}J^2. \tag{14}$$

From Eq. (4), it follows that $\dot\rho_\gamma=0$ if $\rho_\gamma=0$. Therefore trajectories do not cross the 2-dimensional $\rho_\gamma=0$ surface, which is a hyperboloid in the variables H, $\cos(I/2)$, and J. Physical trajectories with $\rho_\gamma>0$ are forced to lie within the volume of the H,I,J phase space bounded by the $\rho_\gamma=0$ surface.

The only critical points of the system (11)–(13) at finite values of H,I,J occur at

(1) $C_{1\pm}$ at $H=\pm H_1$, $I=0\bmod2\pi$, $J=0$; and

(2) C_2 at $H=0$, $I=\pi\bmod2\pi$, $J=0$, where

$$H_1\equiv\sqrt{\frac{2}{3}}m. \tag{15}$$

Both of these points in fact lie on the $\rho_\gamma=0$ surface. Furthermore, this surface intersects the $H=0$ plane only at the isolated points C_2. The $H>0$ and $H<0$ subspaces are thus physically distinct, and the $H<0$ subspace simply corresponds to the time-reversal of the $H>0$ subspace. Therefore we can take $H>0$ without loss of generality.

The pattern of trajectories close to the $\rho_\gamma=0$ surface can be ascertained by continuity to the $\rho_\gamma=0$ solutions, even though the latter are not physical. The $\rho_\gamma=0$ subspace is obtained, for example, by regarding Eq. (14) as a quadratic equation for H, and using the solution to eliminate H, thereby obtaining a 2-dimensional system for I and J given by Eqs. (12) and (13).

We plot the resulting $H>0$ pattern of trajectories in Fig. 1 for values of $I\in[0,2\pi)$. Since the potential, $V(\phi)$, is periodic the same pattern of trajectories repeats itself as we ex-

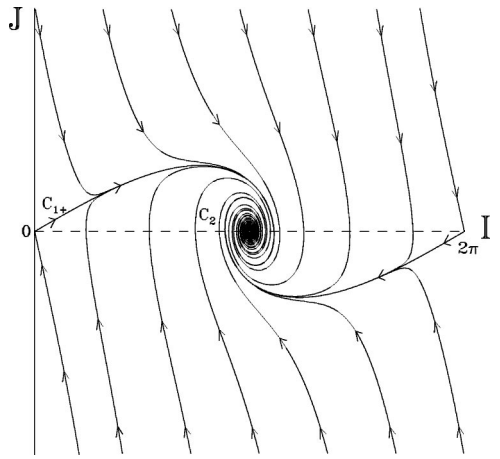

FIG. 1. The projection of the trajectories within the $\rho_\gamma=0$ subspace on the I–J plane for values of $I\in[0,2\pi)$. Within this subspace, C_{1+} is a saddle point and C_2 is a stable spiral.

tend I to $\pm\infty$, with trajectories crossing from one ''cell'' to another at the cell boundaries.

The trajectories which occupy the lower half of Fig. 1 are obtained from those in the upper half by the symmetry $I\rightarrow 2\pi-I$, $J\rightarrow -J$ of the differential equations (11)–(14). Physically this simply corresponds to the scalar field rolling from the maximum to the right of the minimum as opposed to the one to the left.

An analysis of small perturbations about the critical points C_1 and C_2 yield eigenvalues

(1) $\lambda=-\sqrt{6}m\gamma,-\frac{1}{2}m(\sqrt{6}\pm\sqrt{6+4/\mathcal{F}^2})$ at C_{1+},

(2) $\lambda=0,\pm im/\mathcal{F}$ at C_2.

Thus C_{1+} attracts a two-dimensional bunch of trajectories but is a saddle point with respect to trajectories lying in the $\rho_\gamma=0$ surface, as is evident from Fig. 1. The 2-dimensional bunch of trajectories which approach C_{1+} are found to correspond to an inflationary solution with $a\propto\exp(\sqrt{2/3}mt)$ as $t\rightarrow\infty$ and $\phi\rightarrow\text{const}=2n\pi\mathcal{F}$, $n\in$ Z. The possible role of scalar fields with PNGB potentials in driving an inflationary expansion of the early universe has been discussed in [28].

The point C_2 is a degenerate case, in particular with regard to perturbations orthogonal to the $\rho_\gamma=0$ surface (i.e. into the surface $\rho_\gamma>0$ region), for which the eigenvalue is zero. It is a center with respect to the trajectories lying in the $\rho_\gamma=0$ surface, and when perturbations of higher order are considered it becomes a stable spiral point in the $\rho_\gamma=0$ surface as can be seen in Fig. 1. Since there is a degeneracy, however, an alternative choice of phase space variables is desirable. We will defer a discussion of the late time behavior of the solution near C_2 to Sec. II B.

The points $C_{1\pm}$ correspond to models with a scalar field sitting at the maximum of the potential, whereas C_2 corresponds to the scalar field sitting at the bottom of the potential well. The separatrices in Fig. 1 which join $C_{1\pm}$ to C_2 correspond to the field rolling from the maximum to the minimum. It would appear from Fig. 1 that trajectories which spiral into C_2 become arbitrarily close to the separatrix at late times.

The separatrices which join the points $C_{1\pm}$ to points at infinity correspond to solutions for which the scalar field reaches the top of the potential hill as $t\rightarrow\pm\infty$ (e.g., the rightmost trajectory in Fig. 1). Finally, there are also straight line separatrices parallel to the H-axis at each of the points $C_{1\pm}$, extending from $H=\pm H_1$ to infinity, which represent solutions with a static scalar field sitting on top of the potential hill.

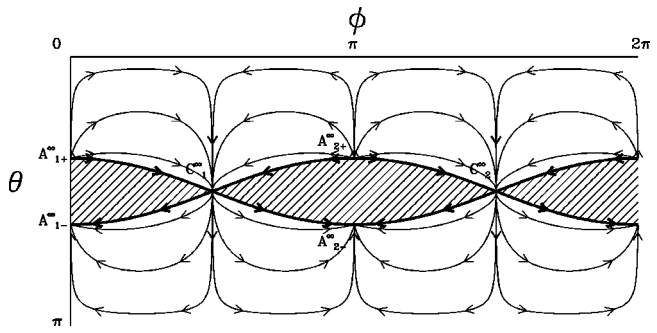

FIG. 2. The projection of trajectories within the sphere at infinity on the ϕ–θ plane. The unphysical region is shaded.

To examine the critical points at infinity it is convenient to transform to spherical polar coordinates r, θ, and ϕ by defining

$$H=r\cos\theta, \tag{16}$$

$$I=r\sin\theta\sin\phi, \tag{17}$$

$$J=r\sin\theta\cos\phi, \tag{18}$$

and to bring the sphere at infinity to a finite distance from the origin by the transformation $r=\rho/(1-\rho)$, $0\leqslant\rho\leqslant 1$ [29].

Although the trajectories on the sphere at infinity do not represent physical cosmologies, it is useful to plot them since the form of the trajectories which lie just within the sphere will be similar. On the sphere $\rho=1$ we find

$$\frac{d\theta}{d\xi}=\sin\theta\cos^2\theta\left[3\sin^2\phi+\frac{2-\gamma}{4}(\tan^2\theta\cos^2\phi-6)\right], \tag{19}$$

$$\frac{d\phi}{d\xi}=\frac{3}{2}\cos\theta\sin 2\phi, \tag{20}$$

where ξ is a new time coordinate defined by $d\xi=rdt$. The resulting integral curves are plotted in Fig. 2. By Eq. (14) the projection of the physical region $\rho_\gamma>0$ onto the sphere at infinity leads to the condition

$$\cot^2\theta>\frac{1}{6}\cos^2\phi. \tag{21}$$

Values of θ and ϕ which violate this inequality lie in the shaded region.

The critical points on the sphere at infinity are

(1) $A^\infty_{1\pm}$, $A^\infty_{2\pm}$: four points at

$(\theta,\phi)\in\{(\pm\tan^{-1}\sqrt{6},0),(\pm\tan^{-1}\sqrt{6},\pi)\}$ or

$H=\pm\infty$, $I/H=0$, and $J/H=\pm\sqrt{6}$.

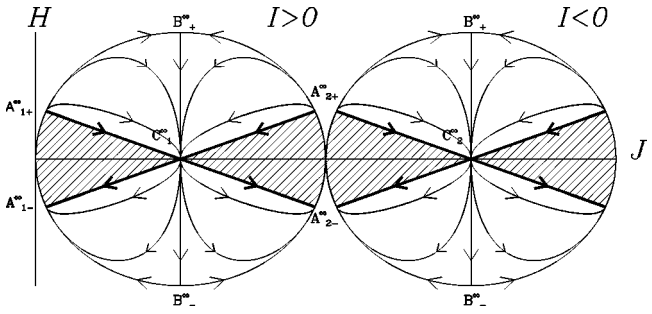

FIG. 3. The projection of trajectories within the $I>0$ and $I<0$ hemispheres at infinity on the H–J plane. The unphysical region is shaded.

(2) $B^{\infty}_{\pm}$: two points at
$H=\pm\infty$, $I/H=0$, and $J/H=0$.

Since the projection onto spherical polar coordinates (16)–(18) is degenerate at the north and south poles $\theta=0,\pi$, these points are excluded from the chart (θ,ϕ) but can be included using an alternative hemispherical projection (see Fig. 3).

(3) $C^{\infty}_{1,2}$: two points at
$(\theta,\phi)\in\{(\pi/2,\pi/2),(\pi/2,3\pi/2)\}$ or
$I=\pm\infty$, $H/I=0$, and $J/I=0$.

An analysis of small perturbations shows that the points $A^{\infty}_{1+,2+}$ are repellors in all directions of the phase space (cf. Table I), while $A^{\infty}_{1-,2-}$ are attractors. This therefore represents the most ''typical'' early behavior of solutions. For $H>0$, $A^{\infty}_{1+,2+}$ correspond to the limit $t\to 0$. We find that $H\sim 1/(3t)$ or $a\propto t^{1/3}$, while $\kappa\phi\sim\pm\sqrt{2/3}\ln t$, for these solutions. The points $A^{\infty}_{1-,2-}$ with $H<0$ represent the time-reversed solutions.

The point B^{∞}_{+} (B^{∞}_{-}) repels (attracts) a 2-dimensional bunch of trajectories traveling to (from) finite values of H,I,J, but is a saddle point with respect to directions on the sphere at infinity. The points are found to correspond to $t\to 0$ with $H\sim 2/(3\gamma t)$ or $a\propto t^{2/3\gamma}$ while $\kappa\phi\propto t^{n}$, $n>0$.

The points $C^{\infty}_{1,2}$ are the projection of the points $C_{1\pm}$ and C_2 into the sphere at infinity. The degenerate eigenvalues simply reflect the degeneracy of the projection.

B^{∞}_{+} acts as a repellor for trajectories with $\dot\phi\simeq 0$, $\phi\simeq$const as $t\to 0$. As shown in Fig. 3, trajectories are driven towards B^{∞}_{+} before they reach C^{∞}_{1}. This is consistent with the property that when H is large ($3H\geqslant m_{\phi}$), the field evolution is over-damped by the expansion, and the field is effectively frozen to its initial value ($\dot\phi\to 0$).

TABLE I. The critical points on the sphere at infinity and their eigenvalues.

Critical points	Eigenvalues (with degeneracies)
$A^{\infty}_{1\pm}, A^{\infty}_{2\pm}$	± 3 (2), $\pm 3(2-\gamma)$
$B^{\infty}_{\pm}$	$\pm\frac{3\gamma}{2}$ (2), $\mp\frac{3}{2}(2-\gamma)$
$C^{\infty}_{1,2}$	0 (3)

B. Energy density variables

In view of the eigenvalue degeneracy encountered above, we can alternatively choose to represent the system by the variables H, x, and y, where

$$x\equiv\frac{\kappa\dot\phi}{\sqrt{6}H}=\frac{J}{\sqrt{6}H}, \tag{22}$$

$$y\equiv\frac{\kappa\sqrt{V}}{\sqrt{3}H}=\frac{\sqrt{2}m\cos(I/2)}{\sqrt{3}H}. \tag{23}$$

These are the same variables used by Copeland, Liddle and Wands [19] for the model with an exponential potential. As above, we will consider $H>0$ only. The field equations then take the form

$$\dot H=-\frac{3}{2}H^2\mu, \tag{24}$$

$$\dot x=\pm\frac{my}{\mathcal{F}}\sqrt{1-\frac{3y^2H^2}{2m^2}}+\frac{3}{2}Hx(\mu-2), \tag{25}$$

$$\dot y=\mp\frac{mx}{\mathcal{F}}\sqrt{1-\frac{3y^2H^2}{2m^2}}+\frac{3}{2}Hy\mu, \tag{26}$$

where

$$\mu(x,y)\equiv\gamma(1-y^2)+(2-\gamma)x^2. \tag{27}$$

We note that in these variables

$$x^2+y^2=\Omega_{\phi} \tag{28}$$

which is why we have adopted the terminology ''energy density variables.'' The physical region of the phase space will be constrained to lie within the cylinder $x^2+y^2\leqslant 1$ since $\Omega_{\phi}\leqslant 1$. In the case of the exponential potential analyzed in Ref. [19], one of the differential equations decoupled, and the dynamics was effectively described by a phase plane with trajectories bounded by the circle $x^2+y^2=1$. In the present case, however, no such simplification arises.

The physical region of phase space is further restricted by the requirement that $y^2\leqslant\frac{2}{3}m^2/H^2\equiv H_1^2/H^2$, which is equivalent to $\cos^2(I/2)\leqslant 1$ in terms of the field variables. For values of $H>H_1$ each $H=$const slice of the cylinder $x^2+y^2\leqslant 1$ is cut off in the y-direction above and below the $y=\pm H_1/H$ lines. Thus the ''fundamental cell'' of the phase space can be considered to be a cylinder for $0\leqslant H\leqslant H_1$, capped by a horn for $H>H_1$, which tapers off to a line segment $-1\leqslant x\leqslant 1$ on the x-axis as $H\to\infty$ (see Fig. 4). In fact, the phase space consists of an infinite number of copies of the fundamental cell of Fig. 4 as a result of the periodic structure of the potential. These cells, $\mathcal{C}_n$, can be labeled by an integer, n, with the variable I lying in the range $2n\pi\leqslant I<2(n+1)\pi$ for each n. For cells with even n the dynamics is described by Eqs. (24)–(27) with the upper sign in Eqs. (25) and (26), while for odd n one must take the lower sign.

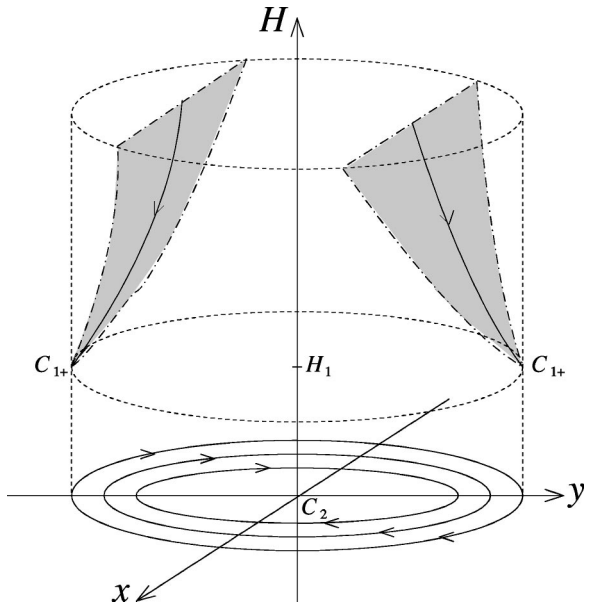

FIG. 4. The ''fundamental cell'' of the phase space in terms of the energy density variables. On the $y^2=H_1^2/H^2$ planes (the shaded planes), only one trajectory is possible as shown. It corresponds to the scalar field lying on top of the potential hill all along. On the $H=0$ plane, the trajectories are concentric circles with center at the origin.

Within the horn portion of each phase space cell the motion of most trajectories is roughly circular in $H=$const slices, in a clockwise sense in even cells, $\mathcal{C}_{2n}$, and anticlockwise in odd cells $\mathcal{C}_{2n+1}$. However, trajectories can cross from one cell to another along the $y=\pm H_1/H$ boundaries of their horns, which correspond to the surfaces $I=2n\pi$ in terms of the field variables. For even cells, $\mathcal{C}_{2n}$, trajectories join the cell $\mathcal{C}_{2n}$ along the $y=H_1/H$ surface and the cell $\mathcal{C}_{2n+1}$ along the $y=-H_1/H$ surface. Below the $H=H_1$ plane solutions cannot cross from one cell to another, but remain confined within the cylinder $x^2+y^2\leqslant 1$.

When $H=0$ we see from Eqs. (24)–(27) that $\dot{x}=my/\mathcal{F}$ and $\dot{y}=-mx/\mathcal{F}$, so that trajectories which lie in the $H=0$ surface are purely concentric circles. Since $\dot{a}=0$ in the $H=0$ plane, these do not represent physically interesting cosmologies, but by continuity the behavior of the trajectories just above the plane will be of a spiral nature.

The origin $H=x=y=0$ is in fact the critical point corresponding to C_2. The nature of the critical point is altered by the change of variables, however. In particular, whereas the eigenvalues for linear perturbations are unchanged, when higher order corrections are considered the point is no longer always an asymptotically stable spiral as was the case in Fig. 1.

Asymptotically as $t\to\infty$ we have

$$x=A(t)\sin\frac{mt}{\mathcal{F}}, \tag{29}$$

$$y=A(t)\cos\frac{mt}{\mathcal{F}}, \tag{30}$$

where the amplitude is governed by the equation

$$\frac{d}{dt}(A^2)=3HA^2\left(\gamma-2\sin^2\frac{mt}{\mathcal{F}}\right)(1-A^2). \tag{31}$$

The nature of C_2 is now found to depend on γ:

(1) If $\gamma<1$ we find that over a cycle the average value of the right-hand side (rhs) of Eq. (31) is negative and A^2 decreases so that C_2 is an asymptotically stable spiral. Furthermore to leading order $H\sim 2/(3\gamma t)$ as $t\to\infty$ or $a\propto t^{2/3\gamma}$ and $A(t)$ is given by

$$A(t)=Bt^{(\gamma-1)/\gamma}\exp\left[-\frac{1}{\gamma}\int_t^\infty\frac{dt'}{t'}\cos\left(\frac{mt'}{\mathcal{F}}\right)\right] \tag{32}$$

with B constant. The late-time attractor has $\Omega_\phi=0$ and $\Omega_\gamma=1$.

(2) If $\gamma>1$ then over a cycle the average value of the rhs of Eq. (31) is positive and A^2 increases until it reaches a limit cycle $A^2=1$, i.e. $x^2+y^2=1$ or $\Omega_\phi=1$, $\Omega_\gamma=0$. In this case

$$A(t)=1-\overline{B}t^{2(1-\gamma)}\exp\left[-2\int_t^\infty\frac{dt'}{t'}\cos\left(\frac{mt'}{\mathcal{F}}\right)\right]. \tag{33}$$

(3) If $\gamma=1$, which corresponds physically to an ordinary matter-dominated universe, then an intermediate situation obtains. Essentially any of the concentric circles in the $H=0$ plane of Fig. 4 can be approached asymptotically giving a universe for which $\Omega_\phi\to\alpha_1$ and $\Omega_\phi\to 1-\alpha_1$ where α_1 is a constant in the range $0<\alpha_1<1$, which depends on the initial conditions and the parameters m and $\mathcal{F}$.

We observe that for all values of γ,

$$\rho_\phi\propto\frac{1}{a^3} \tag{34}$$

at late times. Since $\rho_\gamma\propto a^{-3\gamma}$, the three different late time behaviors can thus be understood as a consequence of the scalar field either decreasing more rapidly than the barotropic fluid ($\gamma<1$), less rapidly ($\gamma>1$), or at the same rate ($\gamma=1$). The scalar field thus eventually dominates if $\gamma>1$, while the barotropic fluid dominates if $\gamma<1$. In the interesting critical case of dust filled models ($\gamma=1$) both the scalar field and ordinary matter are of cosmologically significant density at late times.

One simplification that is often made in studying quintessence models is to assume that at late times the quintessence field obeys an equation of state

$$P_\phi=(\gamma_\phi-1)\rho_\phi \tag{35}$$

with γ_ϕ effectively constant. While such an assumption is justified in the case of models with a slowly varying scalar field, it does not apply in the present case. In particular, since the effective barotropic index of the scalar field is [15]

$$\gamma_\phi=\frac{2x^2}{x^2+y^2}, \tag{36}$$

we see that $\gamma_\phi \simeq \sin^2(mt/\mathcal{F})$ at late times, so that it remains truly variable, varying from 0 to 2 over each cycle. It follows from Eqs. (25)–(28) that the scalar energy density parameter obeys the equation

$$\dot{\Omega}_\phi = 3H\Omega_\phi(1-\Omega_\phi)(\gamma-\gamma_\phi) \tag{37}$$

so that $\dot{\Omega}_\phi \to 0$ as $t \to \infty$, which accords with the late time properties of the solutions observed above.

III. NUMERICAL INTEGRATION

We will now consider the extent to which models based on the PNGB potential are constrained by the latest observational evidence. This work extends the studies previously undertaken by various authors [17,18,23,24,30]. In the most recent analysis, Frieman and Waga have compared the constraints imposed by the high redshift supernovae luminosity distance on the one hand and gravitational lensing bounds on the other [24]. The two measures provide tests which are potentially in opposition. Here we will perform a similar analysis for the PNGB models, but also taking into account the possibility of luminosity evolution which has not been considered in previous studies [18,24]. We are therefore considering the model of the previous section with $\gamma = 1$.

To proceed it is necessary to integrate the equations numerically. To do this we introduce the dimensionless variables

$$u = \frac{J}{H_0} = \frac{\kappa\dot{\phi}}{H_0}, \tag{38}$$

$$v = \Omega_{m0}^{1/3}(1+z), \tag{39}$$

$$w = I = \frac{\kappa\phi}{\mathcal{F}}. \tag{40}$$

where Ω_{m0} is the fractional energy density of matter at the present epoch, t_0. The dynamical system then becomes

$$u' = -3\frac{H}{H_0}u + \frac{m^2}{\mathcal{F}H_0^2}\sin w, \tag{41}$$

$$v' = -v\frac{H}{H_0}, \tag{42}$$

$$w' = \frac{u}{\mathcal{F}}, \tag{43}$$

where the Hubble parameter is defined implicitly according to

$$\frac{H}{H_0} \equiv \left[v^3 + \frac{1}{6}u^2 + \frac{m^2}{3H_0^2}(\cos w + 1) \right]^{1/2}, \tag{44}$$

and prime denotes a derivative with respect to the dimensionless time parameter $\tau \equiv H_0 t$. Only the variable v differs from those used by Frieman and Waga [18]. Our reason for making the choice (39) is that it allows us to integrate the Friedmann equation directly rather than a second order equation (11) which follows from the other equations by virtue of the Bianchi identity. This may possibly lead to better numerical stability since it is not necessary to implement the Friedmann constraint separately.

We begin the integration at initial values of u, v, and w chosen to correspond to initial conditions expected in the early matter-dominated era. The integration proceeds then until the rhs of Eq. (44) is equal to 1, thereby determining the value of the present epoch, t_0, to be the time at which $H = H_0$. We are then also able to determine Ω_{m0}, since according to Eq. (39)

$$\Omega_{m0} = v^3(t_0). \tag{45}$$

The choice of appropriate initial conditions has been previously discussed [23,17,18]. In particular, since the Hubble parameter is large at early times, it effectively acts as a damping term in Eq. (3), driving the scalar field to a state with $\dot{\phi} = 0$ initially, i.e. $u = 0$. We take $v = 1101$ initially, which in view of relation (39) and the fact that $\Omega_{m0} \sim 0.1$–1 corresponds to the early matter dominated era $1100 \leqslant z \lesssim 3000$. Results of the integration do not change significantly if v is altered to values within the same order of magnitude.

The initial value of the scalar field variable, $w_i \equiv w(t_i)$, can lead to some variability in predicted cosmological parameters at the present epoch. A few different values of $w(t_i)$ have been considered by different authors [23,17,18]. However, the only systematic studies of bounds in the M, f parameter space have been performed [18,24] for one particular initial value, $w_i = 1.5$. One must bear in mind that such bounds are also dependent on w_i, the value of which is not greatly restricted. Given that we are starting with $u(t_i) \simeq 0$, so that the kinetic energy of the scalar field is initially negligible, the only physical restriction on the value of w_i comes from the requirement that the scalar field should be sufficiently far from the minimum of the potential, $V(\phi)$, that $\Omega_\phi(t_i)$ is small. Thus will ensure that w_i is consistent with a scalar field that has emerged from the radiation dominated era with Ω_ϕ sufficiently small that is consistent with bounds set by primordial nucleosynthesis, and by structure formation models. This still leaves considerable latitude for the choice of w_i, however.

In Figs. 5 and 6 we display contour plots of $\Omega_{\phi 0}$ and $H_0 t_0$ in the M, f parameter space for two values $w_i = 1.5$ and $w_i = 0.5$. Similar figures have been given by Frieman and Waga [18] in the $w_i = 1.5$ case, although our resolution is somewhat better. As w_i decreases the contour plots do not change significantly in terms of their overall features, but contours with equivalent values shift to lower values of the f parameter. For example, for large values of M the $\Omega_{\phi 0} = 0.7$ contour lies at a value $f \simeq 2.05 \times 10^{18}$ GeV if $w_i = 1.5$, while the same contour lies at $f \simeq 0.94 \times 10^{18}$ GeV if $w_i = 0.5$.

The other principal feature of the plots 5 and 6, which was not commented on in Ref. [18], is the wave-like properties of the contours at larger values of M. These features can be

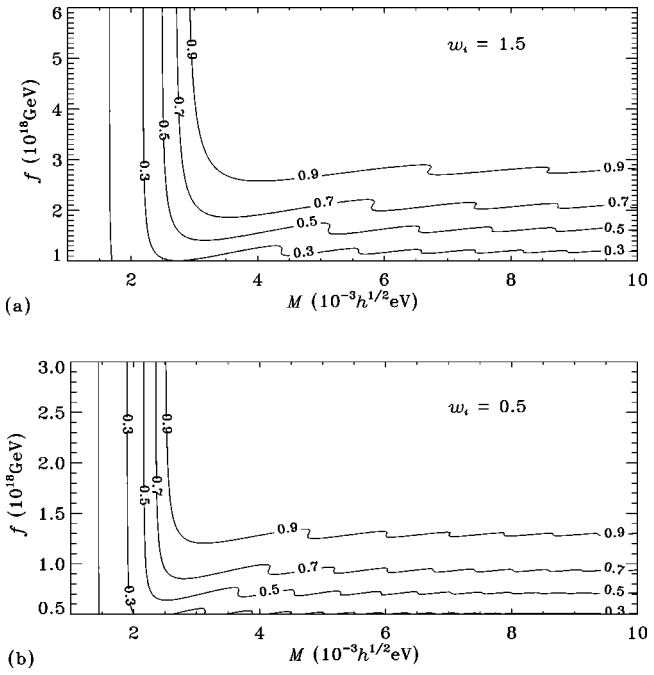

FIG. 5. Contours of $\Omega_{\phi 0}$ in the M,f parameter space for two choices of initial values: (a) $w_i=1.5$; (b) $w_i=0.5$.

readily understood by considering the corresponding plots of the deceleration parameter, q_0, which is defined in terms of $u(t_0)$, $v(t_0)$, and $w(t_0)$ by

$$q_0=\frac{3}{2}v^3(t_0)+\frac{1}{2}u^2(t_0)-1,$$

$$=\frac{1}{2}(1+3\Omega_{\phi 0}-6y_0{}^2). \tag{46}$$

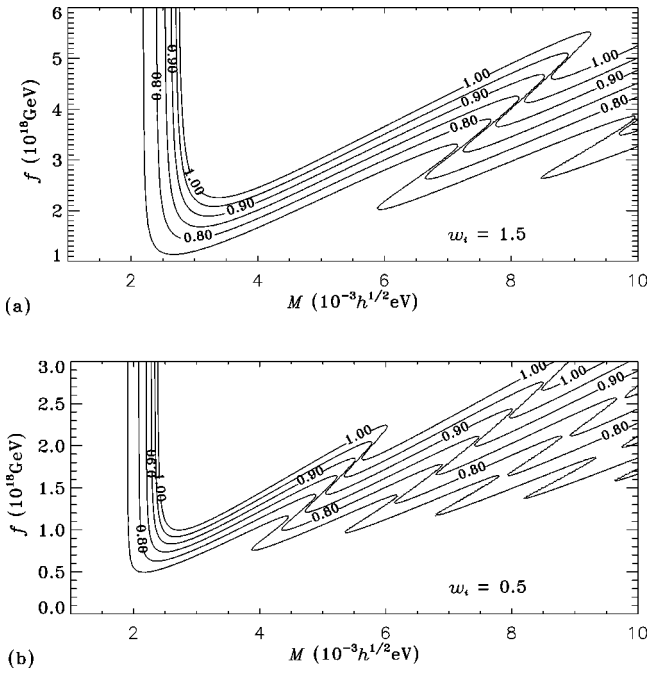

FIG. 6. Contours of H_0t_0 in the M,f parameter space for two choices of initial values: (a) $w_i=1.5$; (b) $w_i=0.5$.

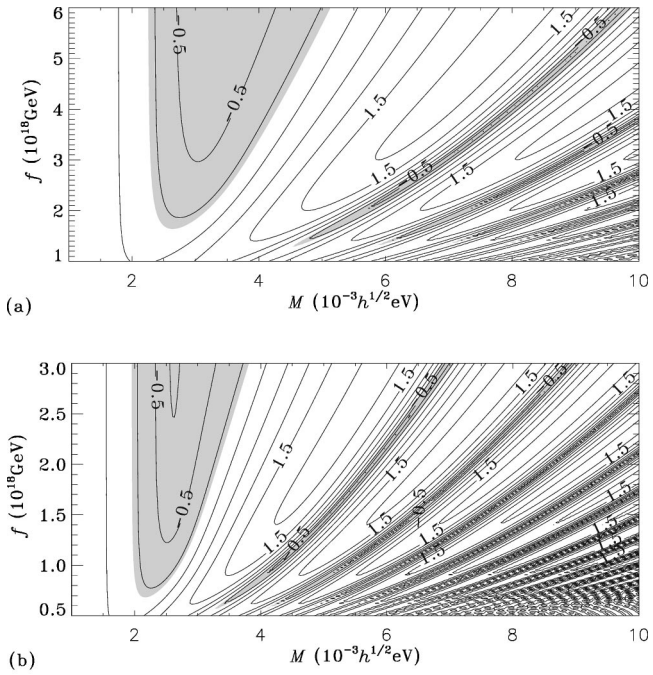

FIG. 7. Contours of q_0 in the M,f parameter space for two choices of initial values: (a) $w_i=1.5$; (b) $w_i=0.5$. Values $q_0<0$, corresponding to a universe whose expansion is accelerating at the present epoch, are shaded.

We display contour plots of q_0 in the M,f parameter space in Fig. 7. Essentially, as M increases for roughly fixed f for sufficiently large M, the value of q_0 oscillates over negative and positive values, from a minimum of $q_0=\frac{1}{2}(1-3\Omega_{\phi 0})$ to a maximum $q_0=\frac{1}{2}(1+3\Omega_{\phi 0})$ about a mean of $q_0=0.5$. This corresponds to the scalar field, ϕ, having undergone more and more oscillations by the time of the present epoch. The minimum value of q_0 is attained when $\dot{\phi}=0$ instantaneously, while the maximum value of q_0 is attained when ϕ is instantaneously passing through the minimum of its potential.

For smaller values of M to the left of the plots the scalar field has only relatively recently become dynamical, whereas for larger values of M, the scalar can already have undergone several oscillations by the time of the present epoch, particularly if f is small. This variation can be understood in terms of the asymptotic period of oscillation of solutions which approach C_1, which by Eqs. (29),(30) is

$$t_a=2\pi\mathcal{F}/m=2\pi f/M^2. \tag{47}$$

The period t_a is shorter for larger M, or for smaller f. Since the final f values plotted in the $w_i=0.5$ case are a factor of 2 smaller than the $w_i=1.5$ case, this also explains why points with the same value of M have undergone more oscillations up to the present epoch for the smaller value of w_i.

The value of H_0t_0 oscillates as M increases for roughly fixed f, according to whether the universe has been accelerating or decelerating in the most recent past, with more rapid variation for parameter values with shorter asymptotic periods, t_a.

The wiggles in the $\Omega_{\phi 0}$ contours are a residual effect of the oscillation of the scalar field as Ω_ϕ settles down to a constant value according to Eq. (37). The variation in the value of $\Omega_{\phi 0}$ with the value of w_i can be understood from the fact that for smaller values of w_i the scalar field begins its evolution in the matter-dominated era closer to the critical points, C_{1+}, corresponding to the maximum of the potential, $V(\phi)$. For fixed M and f the period of quasi-inflationary expansion is therefore longer, and the present value of $\Omega_{\phi 0}$ larger.

IV. CONSTRAINTS FROM HIGH-REDSHIFT TYPE IA SUPERNOVAE

Empirical calibration of the light curve–luminosity relationship of type Ia supernovae provides absolute magnitudes that can be used as distance indicators. Since the luminosity distance of a light source, d_L, is defined by

$$\frac{H_0 d_L}{c} = (z+1)\int_\tau^{\tau_0} (z+1)\, d\tau, \tag{48}$$

with $\tau = H_0 t$, it is convenient to define an additional variable

$$r = -\int_{\tau_0}^{\tau} v\, d\tau, \tag{49}$$

which is proportional to the comoving coordinate distance

$$r_{\text{comoving}} = \frac{r}{H_0 a(t_0) v(t_0)}. \tag{50}$$

We then adjoin a differential equation

$$r' = -v \tag{51}$$

to the differential equations (38)–(40) when performing the numerical integration. In terms of r and v, the luminosity distance is then determined according to

$$\frac{H_0 d_L}{c} = \frac{v}{\Omega_{m0}^{2/3}}(r - r(t_0)), \tag{52}$$

and can be used in the appropriate distance modulus to compare with the supernovae data.

Frieman and Waga [24] have recently considered constraints on PNGB models from supernovae data (without source evolution) using the data set of Riess *et al.* [4]. We will perform a similar analysis, but we will make use of the largest available data set, namely the 60 supernovae published by Perlmutter *et al.* [3] (hereafter P98). Of these, 18 low redshift SNe Ia were discovered and measured in the Calán-Tololo survey [31], and the Supernova Cosmology Project discovered 42 new SNe Ia at redshifts between 0.17 and 0.83. The peak magnitudes of the supernovae are corrected using the ''stretch factor'' light curve fitting method [3,32]. This method is based on fitting a time-stretched version of a single standard template to the observed light curve. The stretch factor is then used to estimate the absolute magnitude.

A. Models without evolution

We use the stretch-luminosity corrected effective B-band peak magnitude in Table 1 of P98 as the absolute magnitude and denote it as m_B^{eff}. The distant modulus of a supernova is defined as $m_B^{\text{eff}} + \mathcal{M}_0$, where $\mathcal{M}_0$ is the fiducial absolute magnitude which has not been given in the literatures. A fitting method to obtain $\mathcal{M}_0$, using only the 18 Calán/Tololo supernovae, can be found in [33,34]. Perlmutter *et al.* [3] calculate the confidence regions by simply integrating over $\mathcal{M}_0$. In this paper, we will perform an analytic marginalization over $\mathcal{M}_0$ and obtain the marginal likelihood.

Similarly to the χ^2 statistic used by Riess *et al.* [4], we define the quadratic form

$$\chi^2(M,f) = \sum_{i=1}^{60} \frac{(m_{B,i}^{\text{eff}} + \mathcal{M}_0 - \mu_i)^2}{\sigma_i^2} \tag{53}$$

where $\mu_i(z_i; H_0, M, f)$ is the predicted distance modulus for model parameters (M,f), and

$$\sigma_i^2 = \sigma_{m_{B,i}^{\text{eff}}}^2 + \left(\frac{5\log \mathrm{e}}{z_i}\sigma_{z,i}\right)^2. \tag{54}$$

The predicted distance modulus is

$$\mu_i = 5\log d_L(z_i) + 25, \tag{55}$$

if the luminosity distance, d_L, as defined by Eq. (48) or Eq. (52), is given in units of Megaparsecs.

In order to perform analytic marginalization over H_0 as well as over $\mathcal{M}_0$ we separate out the H_0 and $\mathcal{M}_0$ dependence from μ_i into a quantity, ν, which we define by

$$\mu_i - \mathcal{M}_0 = g_i + \nu \tag{56}$$

where $g_i(z_i; M, f)$ depends implicitly only on M and f. We then follow the statistical procedures adopted by Drell, Loredo and Wasserman [35,36] and marginalize over ν using a flat prior that is bounded over some range $\Delta\nu$.

The marginal likelihood is

$$\mathcal{L}(M,f) = \frac{1}{\Delta\nu}\int d\nu\, \mathrm{e}^{-\chi^2/2} = \frac{s\sqrt{2\pi}}{\Delta\nu}\mathrm{e}^{-Q/2} \tag{57}$$

where

$$Q(M,f) = \sum_{i=1}^{60} \frac{(m_{B,i}^{\text{eff}} - g_i - \hat{\nu})^2}{\sigma_i^2} \tag{58}$$

and $\hat{\nu}(M,f)$ is the best-fit value of ν given M and f, with conditional uncertainty s. We identify $1/s^2$ as the coefficient of ν^2 in Eq. (53),

$$\frac{1}{s^2} = \sum_{i=1}^{60} \frac{1}{\sigma_i^2}, \tag{59}$$

and $\hat{\nu}$ as s^2 times the coefficient of $-\nu$, viz.

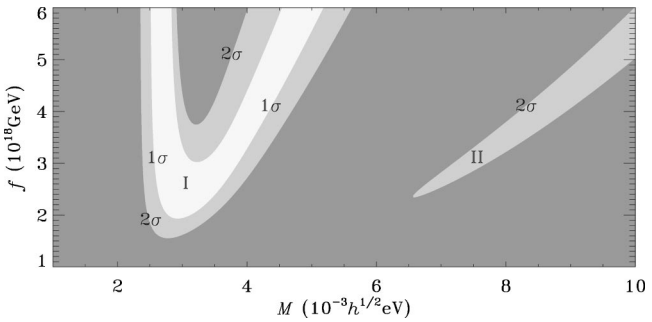

FIG. 8. Confidence limits on M,f parameter values, with $w_i = 1.5$ and no evolution of sources, for the 60 supernovae Ia in the P98 dataset. Parameter values excluded at the 95.4% level are darkly shaded, while those excluded at the 68.3% level are lightly shaded.

$$\hat{\nu}(M,f) = s^2 \sum_{i=1}^{60} \frac{m_{B,i}^{\rm eff} - g_i}{\sigma_i^2}. \tag{60}$$

The quadratic form (58) is what one would obtain by calculating the ''maximum likelihood'' for M and f. Since s is independent of M and f, it follows from Eq. (57) that the marginal likelihood is proportional to the maximum likelihood in this case.

Figure 8 shows the 68.3% and 95.4% joint credible regions for M and f, and is a direct analogue of Fig. 1 of Ref. [24], where a similar analysis was performed on 37 supernovae given in R98. The position of the region of parameters which are included at both the 68.3% and 95.4% confidence levels is broadly similar to that obtained from the R98 data [24]. Although Frieman and Waga [24] did not include the parameter region $M > 0.004h$ eV, we have redone their analysis on the R98 data and find that the parameter values to the right of Fig. 8 which are admitted at the 95.4% level but excluded at the 68.3% for the P98 dataset (labeled region II in Fig. 8), are in fact excluded also at the 95.4% confidence level if the 37 supernovae of the R98 dataset are used. It is possible that this discrepancy has its origin in the different techniques used by Riess *et al.* [4] to determine the distance moduli. Possible systematic discrepancies in the ''stretch factor'' method of P98 versus the ''multi-color light curve'' and ''template fitting methods'' of R98 have been discussed in some detail in Ref. [35].

The importance of the 2σ included parameter region to right of Fig. 8 diminishes, however, if one compares it with Figs. 5 and 6, since it largely corresponds to parameter values with $\Omega_{\phi 0} \gtrsim 0.9$, which can be discounted by dynamical measurements of Ω_{m0}, where $\Omega_{m0} + \Omega_{\phi 0} = 1$. Furthermore, the few allowed values below the $\Omega_{\phi 0} = 0.9$ contour in this part of the parameter space have unacceptably small values for the age of the Universe, $H_0 t_0$.

The parameter region $0.002h$ eV$< M < 0.003h$ eV, which from Fig. 8 is admitted at both the 68.3% and 95.4% levels (labeled region I), by contrast corresponds to acceptable values of both $\Omega_{\phi 0}$ and $H_0 t_0$. Comparing with Fig. 7, we see that this region has $-0.1 \lesssim q_0 \lesssim -0.6$, corresponding to a Universe with a scalar field still in an early stage of rolling down the potential $V(\phi)$. The label I is thus indicative of the fact that the scalar field is rolling down the potential for the first time (from left to right), while in region II the scalar field is rolling down the potential for the second time (from right to left). In region II q_0 is positive—however, it corresponds to parameter values for which there would have been a cosmological acceleration at modest redshifts in the past, e.g., at $z \sim 0.2$, well within the range of the current supernovae dataset.

Since conclusions regarding statistically preferred regions of the parameter space can change if evolution of the sources occurs, we think it is important that this possibility is also examined, as we will now do.

B. Models with evolution

In view of the recent results of Refs. [5,6], supernovae Ia may exhibit some evolutionary behavior, at least as far as their rise times are concerned. A possible evolution in the peak luminosity is therefore a possibility which must be seriously investigated.

In the absence of a detailed physical model to explain precisely how the source peak luminosities vary with redshift, one approach is to assume some particular empirical form for the source evolution, and to examine the consequences. Such an analysis has been recently preformed by Drell, Loredo and Wasserman [35] in the case of Friedmann-Lemaître models with constant vacuum energy. We will undertake an equivalent analysis for the case of PNGB quintessential cosmologies.

Following Drell, Loredo and Wasserman [35] we will assume that the intrinsic luminosities of SNe Ia scale as a power of $1+z$ as a result of evolution. This model introduces a continuous magnitude shift of the form $\beta \ln(1+z)$ to the SNe Ia sample. Equation (53) then becomes

$$\chi^2(M,f) = \sum_{i=1}^{60} \frac{(m_{B,i}^{\rm eff} - g_i - \nu - \beta \ln(1+z_i))^2}{\sigma_i^2}. \tag{61}$$

The parameter β will be assumed to have a Gaussian prior distribution with mean β_0 and standard deviation b. Physically the parameter β_0 represents a redshift-dependent evolution of the peak luminosity of the supernovae sources, which might be expected to arise as a result of the chemical evolution of the environment of the supernovae progenitors as abundances of heavier elements increase with cosmic time. Ultimately, one should hope to account for this evolution with astrophysical modeling of the supernovae explosions [7]. The parameter b would then account for a local variability in the supernovae environments between regions of individual galaxies at the same redshift which are richer or poorer in metals, or with progenitor populations of different ages and masses, etc.

We now have two parameters to marginalize over, ν and β. As in the case of models with no evolution, we will marginalize over ν using a flat prior. We use a Gaussian prior for β with mean β_0 and standard deviation b, so that

$$p(\beta') = \frac{1}{b\sqrt{2\pi}} {\rm e}^{-\beta'^2/2b^2} \tag{62}$$

where

$$\beta' = \beta - \beta_0 . \tag{63}$$

The marginal likelihood is calculated by multiplying the prior (62) by the likelihood resulting from Eq. (57), and integrating over β. The resulting likelihood is

$$\mathcal{L}(M,f) = \frac{1}{\Delta\nu}\int d\beta' \int d\nu\, p(\beta')\mathrm{e}^{-\chi^2/2}$$
$$= \frac{s\bar{\sigma}\sqrt{2\pi}}{\Delta\nu b}\mathrm{e}^{-Q/2} \tag{64}$$

where

$$Q(M,f) = -\frac{\hat{\beta}^2}{\bar{\sigma}^2} + \sum_{i=1}^{60}\frac{(h_i - s^2\mathcal{H})^2}{\sigma_i^2}, \tag{65}$$

$$h_i(z_i;M,f) = m_{B,i}^{\mathrm{eff}} - g_i - \beta_0 \ln(1+z_i), \tag{66}$$

$$\mathcal{H}(M,f) = \sum_{i=1}^{60}\frac{h_i}{\sigma_i^2}. \tag{67}$$

The conditional best-fit of β' is given by

$$\hat{\beta}(M,f) = \bar{\sigma}^2\sum_{i=1}^{60}\frac{h_i[\ln(1+z_i) - s^2\mathcal{G}]}{\sigma_i^2}, \tag{68}$$

$$\mathcal{G} = \sum_{i=1}^{60}\frac{\ln(1+z_i)}{\sigma_i^2}, \tag{69}$$

and the β' uncertainty, $\bar{\sigma}$, is given by

$$\frac{1}{\bar{\sigma}^2} = \frac{1}{b^2} - s^2\mathcal{G}^2 + \sum_{i=1}^{60}\frac{\ln^2(1+z_i)}{\sigma_i^2}. \tag{70}$$

Although $\bar{\sigma}$ is independent of M and f, the marginal likelihood is no longer proportional to the profile likelihood because Q is now given by Eq. (65) rather than by the chi-square type statistic (58).

We have performed a detailed numerical analysis on the P98 dataset, varying β_0, b and w_i. As a result we find a best-fit value of $\beta_0 \equiv \beta^* \simeq 0.414$, for $w_i = 1.5$ in the PNGB models. This would correspond to supernovae being intrinsically dimmer by 0.17 magnitudes at a redshift of $z = 0.5$, which is an effect of the typical order of magnitude being addressed in current attempts to better model the supernova explosions [7]. Furthermore, we find that inclusion of a non-zero variance, b^2, does not alter the prediction of the best-fit value of β_0, although it naturally does lead to a broadening of the areas of parameter space included at the 2σ level. There is relatively little broadening of the region of parameter values included at the 1σ level in the $\beta_0 = \beta^*$ plane, however, which is no doubt a consequence of the steepness of the q_0 contours in Fig. 7(a) in the area corresponding to region II.

In Figs. 9 and 10 we display the joint credible regions for M and f, for two slices through the 3-dimensional (M,f,β_0) parameter space for $w_i = 1.5$: (a) the best-fit case $\beta_0 = \beta^*$; and (b) $\beta_0 = 0$. Analagously, the best-fit case is shown in Fig. 11 for $w_i = 0.2$. We see from Fig. 10 that once the likelihood is normalized relative to β^* no regions remain in the $\beta = 0$ parameter plane which are admitted at the 1σ level when $b = 0$. Furthermore, even when a non-zero standard deviation, b, is included, region II of the (M,f) parameter plane is favored at the 1σ level, in contrast to Fig. 8.

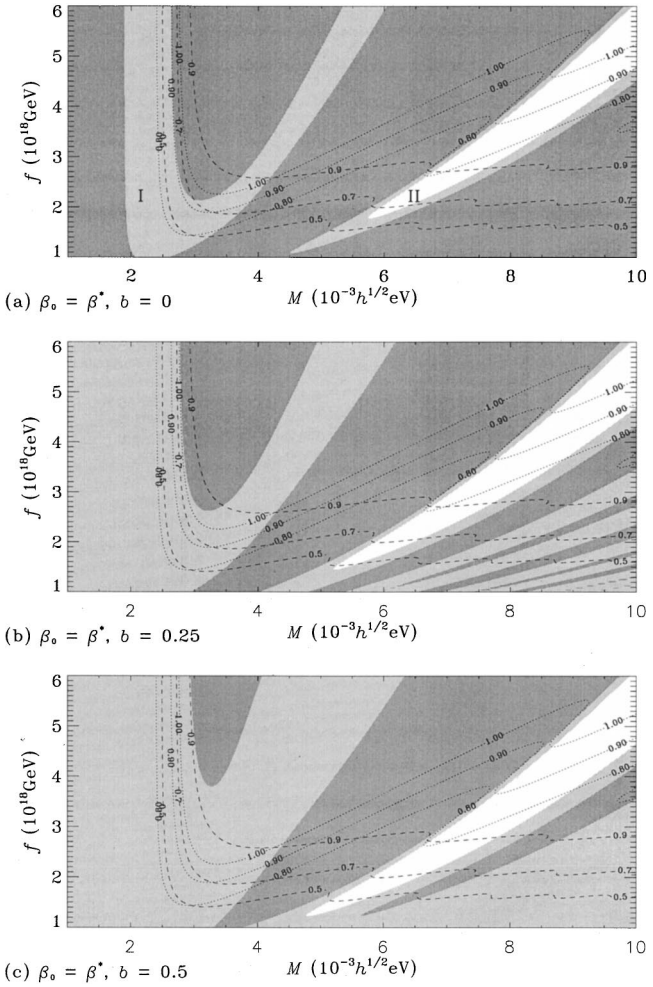

FIG. 9. Confidence limits on M,f parameter values in the best-fit $\beta_0 = 0.414$ slice of the (M,f,β_0) parameter space, with $w_i = 1.5$, for the 60 supernovae Ia in the P98 dataset. Parameter values excluded at the 95.4% level are darkly shaded, while those excluded at the 68.3% level are lightly shaded. For reference, contours of $\Omega_{\phi 0}$ and $H_0 t_0$ are superposed as dashed and dotted lines respectively.

We have also undertaken an analysis of the models with $\beta_0 = 0$ but variable b, similarly to the study of Ref. [35]. In that case, we once again find that region II of the (M,f) parameter plane is admitted at the 1σ level if $b = 0.25$ or $b = 0.5$. The dependence of the value of

$$\bar{Q} \equiv -2\ln(\mathcal{L}\Delta\nu) = Q - 2\ln\left(\frac{\sqrt{2\pi}s\bar{\sigma}}{b}\right) \tag{71}$$

on the value of b is displayed in Fig. 12, for $\beta_0 = 0$ as compared with the best-fit case $\beta_0 = \beta^*$. The quantity $\bar{Q}$ is analo

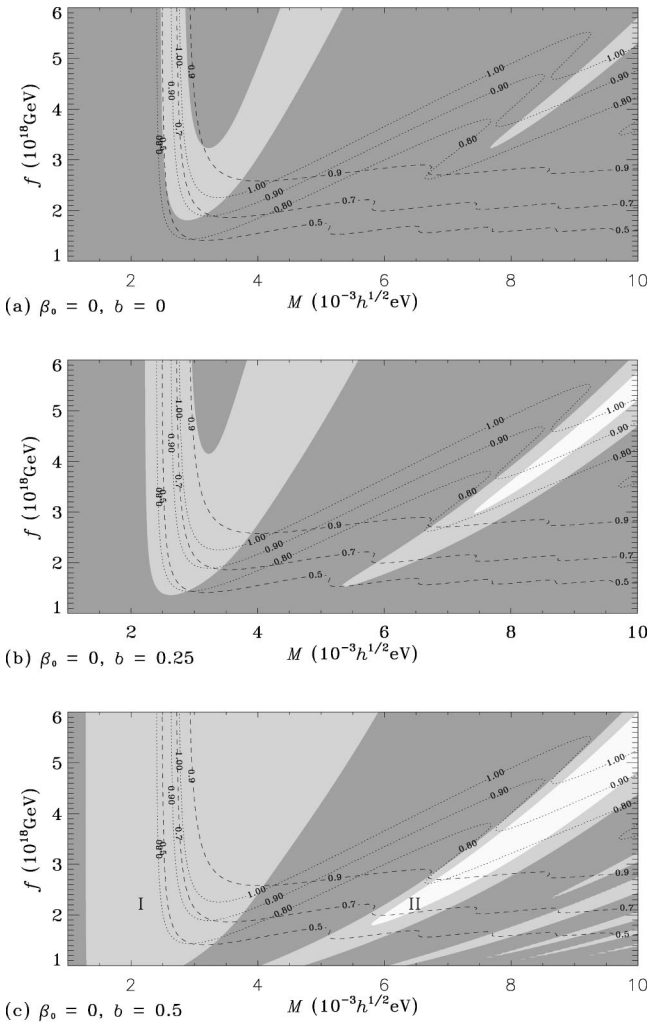

FIG. 10. Confidence limits on M,f parameter values in the $\beta_0=0$ slice of the (M,f,β_0) parameter space relative to a best-fit value $\beta^*\simeq 0.414$, for the 60 supernovae Ia in the P98 dataset, with $w_i=1.5$. Parameter values excluded at the 95.4% level are darkly shaded, while those excluded at the 68.3% level are lightly shaded. For reference, contours of $\Omega_{\phi 0}$ and $H_0 t_0$ are superposed as dashed and dotted lines respectively.

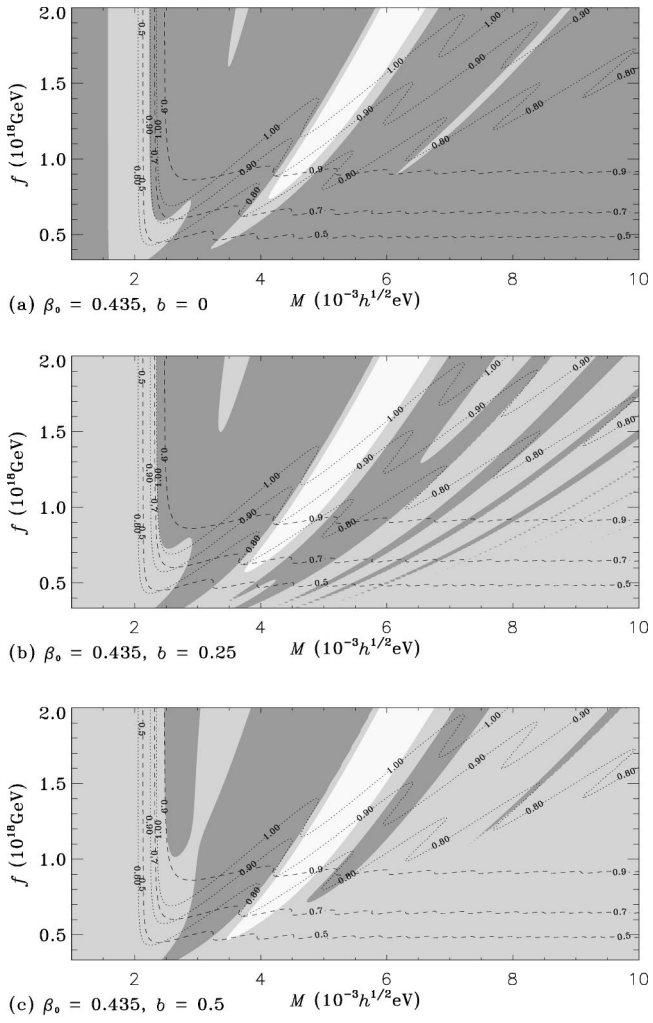

FIG. 11. Confidence limits on M,f parameter values in the $\beta_0=0$ slice of the (M,f,β_0) parameter space relative to a best-fit value $\beta^*\simeq 0.435$, for the 60 supernovae Ia in the P98 dataset, with $w_i=0.2$. Parameter values excluded at the 95.4% level are darkly shaded, while those excluded at the 68.3% level are lightly shaded. For reference, contours of $\Omega_{\phi 0}$ and $H_0 t_0$ are superposed as dashed and dotted lines respectively.

gous to the chi-square statistic of the maximum likelihood method. We see that the $\beta_0=0$ models favor a non-zero value of $b\simeq 0.36$ by a very small margin as compared to the $b=0$ case. For $b=0.36$ the points of greatest likelihood lie mainly in region II, in contrast with the $b=0$ case in Fig. 8.

Varying the initial condition w_i does not appear to affect the best-fit value of β_0 significantly. For $w_i=0.2$ (cf. Fig. 11), for example, the best-fit value was $\beta^*=0.435$, a difference of 5% from the $w_i=1.5$ case. Furthermore, the numerical value of the least value of $\overline{Q}$ was only 0.2% greater in the $w_i=0.2$ case. We have not attempted to find a best-fit value for w_i.

V. CONSTRAINTS FROM LENSING STATISTICS

Gravitational lensing of distant light sources due to the accumulation of matter along the line of sight provide another relatively sensitive constraint on the cosmological models of interest. For cosmology the situation of most interest is the lensing of high luminosity quasars by intervening galaxies. The abundance of multiply imaged quasars and the observed separation of the images to the source puts constraints on the luminosity-redshift relation and hence the model parameters. Basically, if the volume of space to a given redshift is larger then on average one can expect more lensing events. This leads to a statistical test, which has been used to put bounds on Λ [25–27] and to test properties of some decaying Λ or quintessence models [30,37].

Gravitational lensing statistics are useful since they provide a test which potentially provides opposing constraints to those obtained from supernovae magnitude-redshift tests. In particular, in the case of models with a vacuum energy provided by a cosmological constant, the high redshift supernovae have been interpreted as favoring relatively large values of Ω_Λ—Perlmutter *et al.* [3] give a value of $\Omega_\Lambda=0.72$ at 1σ

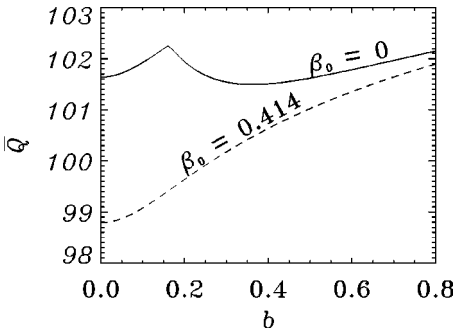

FIG. 12. Variation of the least value of $\bar{Q} = -2\ln(\mathcal{L}\Delta\nu)$ as a function of the standard deviation b of the prior distribution for β_0 for values $\beta_0 = 0$ and $\beta_0 = \beta^*$.

level—whereas the gravitational lensing data lead to upper bounds on Ω_Λ : Kochanek [25] quotes $\Omega_\Lambda < 0.66$ at the 2σ level. A combined likelihood analysis has been performed by various authors [24,27,30].

Gravitational lensing constraints on the PNGB models have been very recently given by Frieman and Waga [24] for $w_i = 1.5$. However, Frieman and Waga considered a more restricted range of parameter space, $M < 0.005h$ eV, since they did not consider the possibility of source evolution in the case of the type Ia supernovae and therefore took values of $M > 0.005h$ eV to be ruled out. We wish to extend the range of M for the gravitational lensing statistics to consider parameter values corresponding to region II in the supernovae constraint graphs, Figs. 8 and 10, so as to compare the constraints from different tests.

We have thus simply followed the calculation described by Waga and Miceli [30], who performed a statistical lensing analysis of optical sources described earlier by Kochanek [25]. They used a total of 862 ($z > 1$) high luminosity quasars plus 5 lenses from seven major optical surveys [38]. (Another alternative not considered here is to analyze data from radio surveys—see, e.g., [26,27].) Undertaking a similar analysis for the increased parameter range, we arrive at Fig. 13, which shows the 68.3% and 95.4% joint credible regions for M and f, for two values of w_i. We refer the reader to Refs. [25,30] for details of the calculation. The only regions of parameter space excluded at the 2σ level turn out to be areas of parameter space for which the deceleration is presently negative (cf., Fig. 7), with the scalar field still commencing its first oscillation at the present epoch.

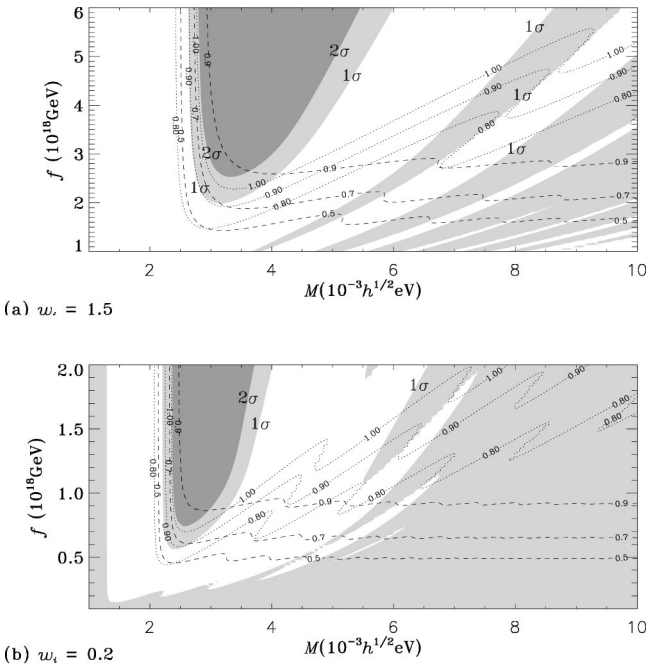

FIG. 13. Confidence limits from gravitational lensing statistics: (a) $w_i = 1.5$; (b) $w_i = 0.2$. Parameter values excluded at the 95.4% level are darkly shaded, while those excluded at the 68.3% level are lightly shaded. For reference, contours of $\Omega_{\phi 0}$ and $H_0 t_0$ are superposed as dashed and dotted lines respectively.

VI. DISCUSSION

Let us now consider the overall implications of the constraints observed above.

First, since empirical models with source evolution do appear to fit the data somewhat better, it would appear that we do have weak evidence for an underlying evolution of the peak luminosity of the type Ia supernovae sources, at least in the context of the PNGB quintessence models. It might be interesting to compare the case of other quintessence models, or the case of a cosmological constant. However, the PNGB model is qualitatively different from such models since its final state corresponds to one in which the ultimate destiny of the universe is to expand at the same rate as a spatially flat Friedmann-Robertson-Walker model, rather than to undergo an accelerated expansion. This is of course precisely why we chose the PNGB models as the basis of our investigation, rather than models in which a late-time accelerated expansion had been built in by hand. If we wish to test the hypothesis that the faintness of the type Ia supernovae is at least partly due to an intrinsic variation of their peak luminosities—which is a very real possibility in view of the results of [5]—then a quintessence model which possesses a variety of possibilities for the present-day variation of the scale factor is probably the best type of model to investigate.

If only supernovae luminosity distances (cf., Fig. 9) and gravitational lensing statistics (cf., Fig. 13) are compared then we see that there is a remarkable concordance between the two tests—region II of Fig. 9 coincides with a region included at even the 1σ level in Fig. 13. This is perhaps not surprising, since in view of Fig. 7 region II corresponds to parameter values for which the present day universe has already undergone almost one complete oscillation of the scalar field about the final critical point C_2 of Fig. 1. It is thus already well on the way towards its asymptotic behavior, which closely resembles that of a standard spatially flat Friedmann-Robertson-Walker model.

Due to the oscillatory behavior, parameter values in region II correspond to models in which there has been a recent cosmological acceleration (e.g., at $z \sim 0.2$), but with a $q(z)$ which changes sign three times over the larger range of redshifts, $0 < z < 4$, in the quasar lensing sample, and therefore differing significantly from Friedmann-Lemaître models over this larger redshift range. Extending the SNe Ia sample

to include objects at redshifts $0.1<z<0.4$ and $z>0.85$ in substantial numbers would greatly improve the ability to decide between models in regions I and II.

There is some cause for concern, however, if we consider the favored values of $\Omega_{\phi 0}$ and $H_0 t_0$. In the case of the models with empirical evolution of the supernovae sources, we always found that the best-fit parameter values occurred at the $\Omega_{\phi 0} \to 1$ boundary of the (M,f) parameter space. The overwhelming evidence of many astronomical observations over the past two decades [39] would tend to indicate that $\Omega_{m0} \simeq 0.2 \pm 0.1$, indicating that a vacuum energy fraction of $\Omega_{\phi 0} \sim 0.7$–0.8 is desirable, and $\Omega_{\phi 0} \lesssim 0.9$ in any case. Although parameter values with $\Omega_{\phi 0} < 0.8$ certainly fall within both the 2σ and 1σ portions of region II of Fig. 9, for all values of b, there are potentially serious problems if we wish to simultaneously obtain large values of $H_0 t_0$. In view of recent estimates of the ages of globular clusters [40], a lower bound of 12 Gyr for the age of the Universe appears to be currently indicated. With $h \simeq 0.65$ this would require $H_0 t_0 \gtrsim 0.8$. For $w_i = 1.5$, parameter values with $H_0 t_0 > 0.8$ coincide with values $\Omega_{\phi 0} \gtrsim 0.9$ in region II, which is phenomenologically problematic.

The tension between the values of $\Omega_{\phi 0}$ and $H_0 t_0$ is somewhat mitigated for lower values of w_i. For $w_i = 0.2$, for example, we see from Figs. 12 and 13(b) that the $\Omega_{\phi 0} = 0.7$ and $H_0 t_0$ meet in region II, and there is a small region of parameters there with $0.7 \lesssim \Omega_{\phi 0} \lesssim 0.9$ and $H_0 t_0 \gtrsim 0.8$, which is also consistent with the other cosmological tests.

Even if the supernovae sources undergo evolution it is clear that parameter values in region I of 8, which are favored in the absence of evolution of peak SNe Ia luminosities, are still included at the 2σ level in the models with evolution, in view of Fig. 9.

Perhaps the most significant aspect of our results is the fact that the mere introduction of an additional dispersion, $b > 0.17$, in the peak luminosities while leaving their mean value fixed (cf. Fig. 11), gives rise to a change in the best-fit region of parameter space from region I to region II. (See [41] for further details.) One would imagine that an increased dispersion is likely to be a feature of many models of source evolution, even if evolutionary effects are of secondary importance. Thus even if the empirical models with non-zero β are somewhat artificial, more sophisticated scenarios could well lead to similar changes in regard to the fitting of cosmological parameters in the PNGB model.

Much tighter bounds on the parameter space of quintessence models, including the present model, will be obtained over the next decade as more supernovae data are collected. What we wish to emphasize, however, is that an effective vacuum energy which is cosmologically significant at the present epoch should not simply be thought of in terms of a ''cosmic acceleration.'' A dynamical vacuum energy with a varying effective equation of state allows for many possibilities for the evolution of the universe, and overly restrictive assumptions, such as equating quintessence to models with a late period of continuous cosmological acceleration, should be avoided. If detailed astrophysical modeling of type Ia supernovae explosions ultimately shows that the dimness of distant supernova events is largely due to evolutionary effects, it does not spell the end for cosmologies with dynamical scalar fields.

ACKNOWLEDGMENTS

We would like to thank Chris Kochanek for supplying us with the gravitational lensing data, Elisa di Pietro, Don Page and Ioav Waga for helpful discussions about various aspects of the paper, and the Australian Research Council for financial support.

[1] R. R. Caldwell, R. Dave, and P. J. Steinhardt, Phys. Rev. Lett. **80**, 1582 (1998); G. Huey, L. Wang, R. Dave, R. R. Caldwell, and P. J. Steinhardt, Phys. Rev. D **59**, 063005 (1999).
[2] S. Perlmutter *et al.*, Nature (London) **391**, 51 (1998); B. P. Schmidt *et al.*, Astrophys. J. **507**, 46 (1998).
[3] S. Perlmutter *et al.*, Astrophys. J. **517**, 565 (1999).
[4] A. G. Riess *et al.*, Astron. J. **116**, 1009 (1998).
[5] A. G. Riess, A. V. Filippenko, W. Li, and B. P. Schmidt, Astron. J. **118**, 2668 (1999).
[6] G. Aldering, R. Knop, and P. Nugent, Astron. J. **119**, 2110 (2000).
[7] D. Arnett, astro-ph/9908169; P. Höflich, K. Nomoto, H. Umeda, and J. C. Wheeler, Astrophys. J. **528**, 590 (2000).
[8] A. G. Riess *et al.*, Astron. J. **118**, 2675 (1999).
[9] P. Steinhardt, in *Critical Problems in Physics*, edited by V. L. Fitch and D. R. Marlow (Princeton University Press, Princeton, 1997).
[10] A. R. Liddle and R. J. Scherrer, Phys. Rev. D **59**, 023509 (1999).
[11] P. J. E. Peebles and B. Ratra, Astrophys. J. Lett. **325**, L17 (1988); B. Ratra and P. J. E. Peebles, Phys. Rev. D **37**, 3406 (1988).
[12] I. Zlatev, L. Wang, and P. J. Steinhardt, Phys. Rev. Lett. **82**, 896 (1999); P. J. Steinhardt, L. Wang, and I. Zlatev, Phys. Rev. D **59**, 123504 (1999); I. Zlatev and P. J. Steinhardt, Phys. Lett. B **459**, 570 (1999).
[13] F. Lucchin and S. Matarrese, Phys. Rev. D **32**, 1316 (1985); Y. Kitada and K. Maeda, Class. Quantum Grav. **10**, 703 (1993).
[14] J. J. Halliwell, Phys. Lett. B **185**, 341 (1987); C. Wetterich, Nucl. Phys. **B302**, 668 (1988); A. B. Burd and J. D. Barrow, *ibid.* **B308**, 929 (1988); A. A. Coley, J. Ibañez, and R. J. van den Hoogen, J. Math. Phys. **38**, 5256 (1997); P. G. Ferreira and M. Joyce, Phys. Rev. D **58**, 023503 (1998); S. C. C. Ng, Phys. Lett. B **485**, 1 (2000).
[15] D. Wands, E. J. Copeland, and A. R. Liddle, Ann. N.Y. Acad. Sci. **688**, 647 (1993).
[16] P. G. Ferreira and M. Joyce, Phys. Rev. Lett. **79**, 4740 (1997).
[17] P. T. P. Viana and A. R. Liddle, Phys. Rev. D **57**, 674 (1998).
[18] J. A. Frieman and I. Waga, Phys. Rev. D **57**, 4642 (1998).
[19] E. J. Copeland, A. R. Liddle, and D. Wands, Phys. Rev. D **57**, 4686 (1998).
[20] A. de la Macorra and G. Piccinelli, Phys. Rev. D **61**, 123503 (2000).

[21] C. T. Hill and G. G. Ross, Nucl. Phys. **B311**, 253 (1988); Phys. Lett. B **203**, 125 (1988).
[22] J. A. Frieman, C. T. Hill, A. Stebbins, and I. Waga, Phys. Rev. Lett. **75**, 2077 (1995).
[23] K. Coble, S. Dodelson, and J. A. Frieman, Phys. Rev. D **55**, 1851 (1997).
[24] I. Waga and J. A. Frieman, Phys. Rev. D **62**, 043521 (2000).
[25] C. S. Kochanek, Astrophys. J. **466**, 638 (1996).
[26] E. E. Falco, C. S. Kochanek, and J. A. Muñoz, Astrophys. J. **494**, 47 (1998).
[27] P. Helbig, Astron. Astrophys. **350**, 1 (1999).
[28] K. Freese, J. A. Frieman, and A. V. Olinto, Phys. Rev. Lett. **65**, 3233 (1990); F. C. Adams, J. R. Bond, K. Freese, J. A. Frieman, and A. V. Olinto, Phys. Rev. D **47**, 426 (1993).
[29] V. A. Belinsky, I. M. Khalatnikov, L. P. Grishchuk, and Ya. B. Zeldovich, Zh. Éksp. Teor. Fiz. **89**, 346 (1985) [Sov. Phys. JETP **62**, 195 (1985)]; D. L. Wiltshire, Phys. Rev. D **36**, 1634 (1987).
[30] I. Waga and A. P. M. R. Miceli, Phys. Rev. D **59**, 103507 (1999).
[31] M. Hamuy *et al.*, Astron. J. **112**, 2408 (1996).
[32] S. Perlmutter *et al.*, Astrophys. J. **483**, 565 (1997).
[33] M. Hamuy, M. M. Phillips, J. Maza, N. B. Suntzeff, R. A. Schommer, and R. Aviles, Astron. J. **109**, 1 (1995).
[34] M. Hamuy, M. M. Philips, R. A. Schommer, N. B. Suntzeff, J. Maza, and R. Aviles, Astron. J. **112**, 2391 (1996).
[35] P. S. Drell, T. J. Loredo, and I. Wasserman, Astrophys. J. **530**, 593 (2000).
[36] Our quantity ν is identical to the quantity γ given in Appendix B of [35].
[37] B. Ratra and A. Quillen, Mon. Not. R. Astron. Soc. **259**, 738 (1992); L. F. Bloomfield Torres and I. Waga, *ibid.* **279**, 712 (1996); V. Silveira and I. Waga, Phys. Rev. D **56**, 4625 (1997); A. R. Cooray and D. Huterer, Astrophys. J. Lett. **513**, L95 (1999).
[38] D. Maoz *et al.*, Astrophys. J. **409**, 28 (1993); D. Crampton, R. D. McClure, and J. M. Fletcher, *ibid.* **392**, 23 (1992); H. K. C. Yee, A. V. Filipenko, and D. H. Tang, Astron. J. **105**, 7 (1993); J. Surdej *et al.*, *ibid.* **105**, 2064 (1993); E. E. Falco, in *Gravitational Lenses in the Universe*, edited by J. Surdej, D. Fraipont-Caro, E. Gosset, S. Refsdal, and M. Remy (Univ. of Liège, Liège, 1994), p. 127; C. S. Kochanek, E. E. Falco, and R. Shild, Astrophys. J. **452**, 109 (1995); A. O. Jaunsen, M. Jablonski, B. R. Pettersen, and R. Stabell, Astron. Astrophys. **300**, 323 (1995).
[39] See, for example, J. P. Ostriker and P. J. Steinhardt, Nature (London) **377**, 600 (1995), and references therein.
[40] L. M. Krauss, Phys. Rep. **333**, 33 (2000).
[41] D. L. Wiltshire, astro-ph/0010443.