J. Fluid Mech. (2001), *vol.* 445, *pp.* 93–120.
DOI: 10.1017/S002211200100550X Printed in the United Kingdom

Nonlinear theory of geostrophic adjustment. Part 1. Rotating shallow-water model

By G. M. REZNIK[1], V. ZEITLIN[2]† AND M. BEN JELLOUL[2]

[1]P. P. Shirshov Institute of Oceanology, Moscow, Russia

[2]LMD, Ecole Normale Supérieure, 24, rue Lhomond, 75005 Paris Cedex 05, France

(Received 3 April 2000 and in revised form 26 April 2001)

We develop a theory of nonlinear geostrophic adjustment of arbitrary localized (i.e. finite-energy) disturbances in the framework of the non-dissipative rotating shallow-water dynamics. The only assumptions made are the well-defined scale of disturbance and the smallness of the Rossby number *Ro*. By systematically using the multi-time-scale perturbation expansions in Rossby number it is shown that the resulting field is split in a unique way into slow and fast components evolving with characteristic time scales f_0^{-1} and $(f_0 Ro)^{-1}$ respectively, where f_0 is the Coriolis parameter. The slow component is not influenced by the fast one and remains close to the geostrophic balance. The algorithm of its initialization readily follows by construction.

The scenario of adjustment depends on the characteristic scale and/or initial relative elevation of the free surface $\Delta H/H_0$, where ΔH and H_0 are typical values of the initial elevation and the mean depth, respectively. For small relative elevations ($\Delta H/H_0 = O(Ro)$) the evolution of the slow motion is governed by the well-known quasi-geostrophic potential vorticity equation for times $t \leqslant (f_0 Ro)^{-1}$. We find modifications to this equation for longer times $t \leqslant (f_0\, Ro^2)^{-1}$. The fast component consists mainly of linear inertia–gravity waves rapidly propagating outward from the initial disturbance.

For large relative elevations ($\Delta H/H_0 \gg Ro$) the slow field is governed by the frontal geostrophic dynamics equation. The fast component in this case is a spatially localized packet of inertial oscillations coupled to the slow component of the flow. Its envelope experiences slow modulation and obeys a Schrödinger-type modulation equation describing advection and dispersion of the packet. A case of intermediate elevation is also considered.

1. Introduction

It is well known that synoptic-scale disturbances in the atmosphere and ocean satisfy, approximately, the conditions of geostrophic balance and it is generally accepted that any sufficiently large-scale perturbation has a tendency to shed its 'redundant' part in a form of rapidly propagating (gravity) waves in order to recover the state of geostrophic equilibrium. This process is called geostrophic (or Rossby, after the pioneering work of Rossby 1938) adjustment. Confirmation of the above-described scenario by *linear* analysis has existed for a long time (cf. Obukhov 1949; Monin & Obukhov 1958; and e.g. the textbook by Gill 1982). Much less is known theoretically about *nonlinear* geostrophic adjustment which is the subject of the present paper. The first step in the nonlinear approach was made by Obukhov in his classical

† Author to whom correspondence should be addressed: zeitlin@lmd.ens.fr

paper (Obukhov 1949). The lowest-order nonlinear corrections were discussed in the review by Blumen (1972) (where many earlier references may be found) and, more recently by Dewar & Killworth (1995). However, full and consistent analysis is still lacking, to our knowledge.

In the present paper we report a study of the nonlinear geostrophic adjustment process of spatially localized perturbations in an unbounded domain in the simplest possible framework of the rotating shallow-water (RSW) or equivalent barotropic model. Application of the same ideas to nonlinear adjustment in the multi-layer and continuously stratified models will be presented in Part 2 of this work. Nonlinear geostrophic adjustment has attracted much attention in the context of (spontaneous) frontogenesis the (e.g. Blumen & Wu 1995 and references therein). This problem, when the spatial locality of the initial perturbation in one direction is relaxed and vertical boundaries are present, will be addressed in Part 3. Below we shall limit our discussion and bibliography exclusively to the RSW situation.

Taken to its full extent the problem of nonlinear geostrophic adjustment largely overlaps with the problem of balanced motion or, more generally, that of the slow manifold and slow–fast variable splitting in the atmosphere and ocean dynamics, in general, and in the RSW model, in particular. This problem has been under active study in recent years (in the context close to the present work see Medvedev 1999, and e.g. Warn *et al.* 1995 which contains most of the essential references; Embid & Majda 1996; Babin, Mahalov & Nicolaenko 1998*a*, *b*; and on the practical importance of balance and splitting for numerical weather prediction see the recent review by Cullen 1998). Formulated in simple terms the fundamental question is about possibility of splitting an arbitrary motion into a slow and a fast component in such a way that the slow component will be not influenced by the fast one for long enough times. The approach we chose in the present paper, although close in spirit to the recent work of Embid & Majda (1996) and Babin *et al.* (1998*a*, *b*) and also based on the classical method of nonlinear averaging, in the theory of dynamical systems, differs, nevertheless, from that in several ways:

(1) *Technical application of the method*: unlike the above-mentioned authors who use direct time averaging we apply the multi-scale asymptotic expansions technique (more familiar to practitioners in fluid mechanics) and average at each order of the perturbation theory.

(2) *Initial and boundary conditions*: we always take the geostrophic adjustment context by choosing spatially localized finite-energy initial disturbances having a broad enough continuous spectrum and evolving in the whole space, in contrast with the standard balance problem setting corresponding to a finite (periodic) domain and modal structure (see below on the difference between the discrete and continuous spectrum cases).

(3) *Parameter range*: we explore not only the standard quasi-geostrophic regime, but also nonlinear geostrophic and frontal geostrophic regimes corresponding to larger-scale/stronger nonlinearity disturbances, for the balanced part of the flow.

The plan of the paper is as follows: In §2 we recall the basic features of the RSW model, introduce characteristic parameters and scales and briefly sketch the properties of known asymptotic regimes for slow motions. In §3 we analyse the nonlinear geostrophic adjustment in the standard quasi-geostrophic (QG) regime on the f-plane. We then repeat the same kind of analysis in the frontal geostrophic and nonlinear geostrophic regimes in §4. In §5 we comment on the changes to be introduced by moving to the β-plane approximation. Summary and conclusions are presented in §6. Appendix A contains the technicalities of stationary-phase calculations necessary for the analysis of resonances in perturbation theory. The details of the intermediate geostrophic regime calculations may be found in Appendix B.

2. Preliminaries

As is well known the RSW model consists of the horizontal momentum and mass conservation equations for a thin free-surface fluid layer under the influence of the Coriolis force and gravity on the rotating plane (x, y)

$$\partial_t \boldsymbol{v} + \boldsymbol{v} \cdot \nabla \boldsymbol{v} + f\hat{\boldsymbol{z}} \wedge \boldsymbol{u} + g\nabla H = 0, \tag{2.1}$$

$$\partial_t H + \nabla \cdot (\boldsymbol{v} H) = 0, \tag{2.2}$$

where $\boldsymbol{v} = (u(x, y, t), v(x, y, t))$ is the two-dimensional velocity field, $H = H_0 + h(x, y, t)$ is the free-surface elevation with the rest state corresponding to constant H_0, f is the Coriolis parameter which is equal to f_0 in the f-plane approximation (to be adopted below unless otherwise stated) and to $f_0 + \beta y$ in the β-plane approximation, g is acceleration due to gravity (g becomes reduced gravity in oceanic applications and H should be replaced by the geopotential height in the equivalent barotropic model in meteorology). Here and below $\partial^n_{abc\ldots}$ denotes the nth partial derivative with respect to $a, b, c, \ldots$, $\nabla = (\partial_x, \partial_y)$ and $\hat{\boldsymbol{z}}$ is the vertical unit vector.

The potential vorticity anomaly which is called simply PV below is defined as

$$q = \frac{\zeta + f}{H} - \frac{f_0}{H_0}, \tag{2.3}$$

where $\zeta = \partial_x v - \partial_y u$ is relative vorticity. The PV conservation equation

$$\partial_t q + \boldsymbol{v} \cdot \nabla q = 0 \tag{2.4}$$

follows from (2.1), (2.2) and will be frequently used as an alternative to (2.2).

In what follows we shall concentrate on the initial-value problem for equations (2.1), (2.2) with the fields $\boldsymbol{v}, h$ satisfying localized initial conditions at $t = 0$:

$$u = u_I(x, y), \qquad v = v_I(x, y), \qquad h = h_I(x, y), \tag{2.5}$$

which are assumed to decay rapidly at infinity and, thus, have finite energy. In this case PV decays at infinity as well. In addition, we will always presume that initial conditions are sufficiently smooth to render all the manipulations below well-defined.

The problem of *linear* adjustment of (2.5) according to (2.1), (2.2), (2.4) has a well-known solution (e.g. Obukhov 1949)

$$\boldsymbol{v} = \boldsymbol{v}^{(g)} + \boldsymbol{v}^{(a)}, \qquad h = h^{(g)} + h^{(a)}, \tag{2.6}$$

where the fields $\boldsymbol{v}^{(g)}, h^{(g)}$ correspond to the geostrophic part which is time-independent and is determined by the initial linearized PV:

$$\nabla^2 h^{(g)} - \frac{1}{R_d^2} h^{(g)} = \frac{f_0}{g} \zeta_I - \frac{1}{R_d^2} h_I, \tag{2.7}$$

$$\boldsymbol{v}^{(g)} = \frac{g}{f} \hat{\boldsymbol{z}} \wedge \nabla h^{(g)}, \tag{2.8}$$

where the initial relative vorticity is $\zeta_I = \hat{\boldsymbol{z}} \cdot \nabla \wedge \boldsymbol{v}_I$ and the Rossby deformation radius is defined as

$$R_d = \frac{\sqrt{g H_0}}{f_0}. \tag{2.9}$$

The linearized PV of the ageostrophic part $\boldsymbol{v}^{(a)}, h^{(a)}$ is identically zero.

The ageostrophic field consists of inertia–gravity waves propagating out of the initial disturbance. Their dispersion relation is

$$\omega = \pm\Omega_{\boldsymbol{k}} = \pm\sqrt{gH_0\boldsymbol{k}^2 + f_0^2}, \tag{2.10}$$

where ω is the wavefrequency and $\boldsymbol{k}$ is wavenumber. Due to the dispersive character of the waves the ageostrophic field decays with increasing time at any fixed space point and the solution tends to the geostrophic state (2.7) as $t \to \infty$, x, y fixed. Obviously, the geostrophic part vanishes if the initial linearized PV is zero, and, vice versa, the radiation field being absent for geostrophically balanced initial conditions.

Taking the nonlinearity into account results in PV changes in time (cf. (2.4)) at the advective time scale $T_a = L/U$, equal to the time required for a fluid particle moving with a typical velocity U to travel over the typical scale L of the initial disturbance. In what follows we assume the advective time scale to be much larger than the typical time scale f_0^{-1} of the inertia–gravity waves. Moreover, from now on we suppose that initial data may be described by a single characteristic scale, L, which allows us to limit ourselves to multiple time but single space-scale asymptotic expansions in what follows. The non-dimensional parameters governing the RSW dynamics are the Rossby and Burger numbers

$$Ro \equiv \epsilon = \frac{U}{f_0L}, \qquad Bu \equiv s = \frac{R_d^2}{L^2}, \tag{2.11}$$

and the nonlinearity parameter $\lambda = \Delta H/H_0$ (ΔH is the characteristic free-surface displacement). Therefore, the condition of time-scale separation is written as

$$(T_af_0)^{-1} = \frac{U}{f_0L} = \epsilon = o(1) \tag{2.12}$$

which means that, whatever the values of the other parameters, the Rossby number is small. Both the fast changes due to inertia–gravity wave activity and the slow changes of PV are present in the evolution of the initial disturbance and the problem of *nonlinear* adjustment is to determine their mutual influence.

The fast and the slow processes in the RSW context have been studied separately in numerous works, most, for obvious reasons, concentrating on slow motions.

On the one hand, purely fast motion may be extracted by imposing a zero PV initial condition. By virtue of its Lagrangian conservation PV then remains equal to zero for all time. This procedure was used by Fal'kovich (1992) and Fal'kovich & Medvedev (1992). On the other hand, an effective filtering of the fast inertia–gravity waves may be achieved, following the classical papers of Charney (1948) and Obukhov (1949) by using exclusively the scale $t_1 = (f_0\epsilon)^{-1}$. Following this prescription a small parameter – Rossby number – appears in front of advective derivatives on the left-hand side of the non-dimensionalized equation (2.1) and, thus, the geostrophic equilibrium

$$f\hat{\boldsymbol{z}} \wedge \boldsymbol{v} + g\nabla H = 0 \tag{2.13}$$

always results in the leading order in ϵ, provided

$$\lambda s = O(\epsilon). \tag{2.14}$$

Relation (2.14) constitutes, together with the smallness of the Rossby number, the necessary conditions for geostrophic balance and shows (Charney & Flierl 1981; Romanova & Zeitlin 1984; Williams & Yamagata 1984; Stegner & Zeitlin 1995) that different dynamical regimes are possible for different ratios of the Rossby number to the characteristic nonlinearity. (For the experimental confirmation of this see Stegner & Zeitlin 1998.) The standard quasi-geostrophic (QG) equations widely used for synoptic motions in the atmosphere and ocean (e.g. Holton 1979; Pedlosky 1982) correspond to the case $\lambda \sim \epsilon \ll 1$ and, hence, to $L \sim R_d$. The frontal geostrophic or frontal dynamics (FD) regime (Cushman-Roisin 1986) is characterized by strong elevations $\lambda = O(1)$ and, consequently, by large scales $L \gg R_d$, according to (2.14). Finally, the intermediate regime called nonlinear geostrophic (NLQG) in (Stegner & Zeitlin 1995) corresponds to $\epsilon \ll \lambda \ll 1$ and to scales large but smaller than in FD.

The geostrophic balance, valid at the zeroth order in ϵ, allows velocity to be expressed in terms of the non-dimensional free-surface elevation, h, where $H = H_0 + \Delta H h = H_0(1 + \lambda h)$ (here and below we keep, for simplicity, the same notation for dimensional and non-dimensional quantities). Plugging these expressions in (2.3), (2.4) results, at the leading order in ϵ, in the following slow-motion equations for the above regimes on the f-plane (cf. Stegner & Zeitlin 1995):

1. QG

$$\partial_t h - \partial_t \nabla^2 h - J(h, \nabla^2 h) = 0; \tag{2.15}$$

2. NLQG

$$\partial_t h - J(h, \nabla^2 h) = 0\,; \tag{2.16}$$

3. FD

$$\partial_t h - J\left(\frac{(1+h)^2}{2}, \nabla^2 h\right) - J\left(h, \frac{(\nabla h)^2}{2}\right) = 0. \tag{2.17}$$

Here and below all parameters of order one are taken to be equal to unity for simplicity of notation and $J(A, B)$ denotes the Jacobian of two functions A, B. In all three cases the balanced equation expresses the PV conservation with PV being calculated at the leading order of the corresponding perturbation theory, cf. (2.20) below.

An essential question is, then, what are the corrections to these equations due to the fast motion (waves) or, in the context of the geostrophic adjustment, what has the dynamics described by the above equations to do with the evolution of a small Rossby number disturbance in the framework of the full RSW equations. The answer to this question will be given in what follows. It will be based on the systematic application of the standard multiple-time-scale perturbation theory which may be briefly summarized as follows:

Non-dimensionalization with the rapid time scale t of equations (2.1), (2.2), (2.4), respectively, gives (under the hypothesis that (2.14) holds)

$$\partial_t \boldsymbol{v} + \epsilon(\boldsymbol{v} \cdot \nabla \boldsymbol{v}) + \hat{\boldsymbol{z}} \wedge \boldsymbol{v} + \nabla h = 0, \tag{2.18}$$

$$\lambda \partial_t h + \epsilon(1 + \lambda h)\nabla \cdot \boldsymbol{v} + \lambda \epsilon \boldsymbol{v} \cdot \nabla h = 0, \tag{2.19}$$

$$\partial_t q + \epsilon \boldsymbol{v} \cdot \nabla q = 0, \qquad q = \frac{\epsilon \zeta - \lambda h}{1 + \lambda h}. \tag{2.20}$$

It is readily seen from these equations that ϵ controls the nonlinearity. Hence, by expanding all variables in asymptotic series in ϵ (in the NLQG regime one should use λ as the largest of the small parameters for this purpose) and introducing, simultaneously, slower times $t_1, t_2, \ldots$ one obtains at each order (except for the first one) of the perturbation theory an inhomogeneous linear problem with the right-hand side expressed in terms of the lower-order solution. At the zeroth order in the small parameter one obtains a linear system with the spectrum described earlier and solutions consisting of waves oscillating in time and zero-mode PV. The time-independence of the PV solution in this approximation means that it may depend on t_1 at most. The consistency of the perturbation theory requires that any resonance on the right-hand side of the higher-order equations should be compensated by a proper slow-time evolution of dynamical variables. At each order of the perturbation theory one, thus, obtains a rapidly evolving and, possibly, slow modulated wave part of the flow and a slow evolving vortex part. The question then is how these two components interact with each other, a perfect splitting corresponding to the absence of all interactions or, at least, to the absence of the influence (drag) of the fast motion upon the evolution of the slow one which is supposedly still given by the balanced equations (2.15), (2.16), (2.17). Note that while dealing with the system (2.18)–(2.20) one should always distinguish among different dynamical regimes depending on the ratio of the two basic parameters ϵ and λ which influences not only behaviour of the slow balanced component, but also that of waves. In non-dimensional form the dispersion relation (2.10) is

$$\omega = \sqrt{s\boldsymbol{k}^2 + 1} \tag{2.21}$$

and it is clear that under the FD scaling ($s \to 0$) the inertia–gravity waves degenerate and become (almost) non-propagating inertial oscillations – a fact which will play a crucial rôle in § 4 below.

3. Nonlinear geostrophic adjustment in the QG regime

3.1. *The lowest order of the perturbation theory*

Applying the approach described above we introduce a hierarchy of time scales $t_n = \epsilon^n t$ and develop all variables in formal asymptotic series as follows:

$$\left.\begin{aligned} \boldsymbol{v} &= \boldsymbol{v}_0(x, y; t, t_1, t_2, \ldots) + \epsilon \boldsymbol{v}_1(x, y; t, t_1, t_2, \ldots) + \cdots, \\ h &= h_0(x, y; t, t_1, t_2, \ldots) + \epsilon h_1(x, y; t, t_1, t_2, \ldots) + \cdots. \end{aligned}\right\} \tag{3.1}$$

At the lowest order we have from (2.18)–(2.20)

$$\partial_t \boldsymbol{v}_0 + \hat{\boldsymbol{z}} \wedge \boldsymbol{v}_0 = -\nabla h_0, \tag{3.2}$$

$$\partial_t(\zeta_0 - h_0) = 0, \tag{3.3}$$

where $\zeta_0 = \hat{\boldsymbol{z}} \cdot \nabla \wedge \boldsymbol{v}_0$, with initial conditions

$$u_0|_{t=0} = u_I, \qquad v_0|_{t=0} = v_I, \qquad h_0|_{t=0} = h_I; \tag{3.4}$$

here and below it is supposed that initial data have no dependence on ϵ. It is convenient to rewrite (3.2) in terms of relative vorticity ζ and divergence D (perturbative expansions for these are also presumed, following from (3.1)):

$$\partial_t \zeta_0 + D_0 = 0, \tag{3.5}$$

$$\partial_t D_0 - \zeta_0 = -\nabla^2 h_0. \tag{3.6}$$

Equation (3.3) may be immediately integrated in fast time giving

$$\zeta_0 - h_0 = \Pi_0, \tag{3.7}$$

where Π_0 is an as yet unknown function of coordinates and slow times. Eliminating ζ_0 and D_0 from (3.5)–(3.7) results in the inhomogeneous linear equation for h_0

$$-\frac{\partial^2 h_0}{\partial t^2} - h_0 + \nabla^2 h_0 = \Pi_0(x, y; t_1, t_2, \ldots). \tag{3.8}$$

Solution of this equation may be written as a combination of fast $\tilde{h}_0$ and slow (zero mode) $\bar{h}_0$ components (in what follows we will always use the overbar notation for the slow and the tilde notation for the fast components of motion):

$$h_0 = \tilde{h}_0(x, y; t, \ldots) + \bar{h}_0(x, y; t_1, \ldots) \tag{3.9}$$

satisfying the equations

$$-\frac{\partial^2 \tilde{h}_0}{\partial t^2} - \tilde{h}_0 + \nabla^2 \tilde{h}_0 = 0, \tag{3.10}$$

$$-\bar{h}_0 + \nabla^2 \bar{h}_0 = \Pi_0 \tag{3.11}$$

(Klein–Gordon and inhomogeneous Helmholtz equations, respectively). Note that Π_0 corresponds to the balanced quasi-geostrophic PV built from the slow component $\bar{h}_0$.

We, thus, get a fast–slow splitting at the level of the equations of motion. To render the procedure consistent, it should also be accomplished at the level of initial conditions (the initialization problem). At initial moment $t = t_1 = t_2 = \ldots = 0$ one has from the definition of Π_0 (3.7) that

$$\Pi_0(x, y; 0) = \partial_x v_I - \partial_y u_I - h_I \equiv \Pi_I(x, y). \tag{3.12}$$

This allows an initial value $\bar{h}_{0I}$ of $\bar{h}_0$ to be found by inverting the 'screened' Laplacian operator in

$$-\bar{h}_{0I} + \nabla^2 \bar{h}_{0I} = \Pi_I \tag{3.13}$$

(our choice of decaying boundary conditions on the plane guarantees that the inversion problem is solvable in a unique way). In turn, knowing $\bar{h}_{0I}$ one obtains the initial condition for $\tilde{h}_0$:

$$\tilde{h}_{0I} = h_I - \bar{h}_{0I}. \tag{3.14}$$

The second initial condition for $\tilde{h}_0$ follows from equations (3.5)–(3.7):

$$\partial_t \tilde{h}_0|_{t=0} = -D_I \equiv \partial_x u_I + \partial_y v_I. \tag{3.15}$$

The Klein–Gordon (KG) equation (3.10) together with initial conditions (3.14), (3.15) form a closed problem for $\tilde{h}_0$.

A similar decomposition is used for the velocity field:

$$\boldsymbol{v}_0 = \tilde{\boldsymbol{v}}_0(x, y; t, \ldots) + \bar{\boldsymbol{v}}_0(x, y; t_1, \ldots), \tag{3.16}$$

where the slow components satisfy the geostrophic balance

$$\bar{\boldsymbol{v}}_0 = \hat{\boldsymbol{z}} \wedge \nabla \bar{h}_0 \tag{3.17}$$

and the fast ones obey

$$\partial_t \tilde{\boldsymbol{v}}_0 + \hat{\boldsymbol{z}} \wedge \tilde{\boldsymbol{v}}_0 = -\nabla \tilde{h}_0 \tag{3.18}$$

with initial conditions

$$\tilde{u}_{0I} = u_I - \bar{u}_{0I}, \qquad \tilde{v}_{0I} = v_I - \bar{v}_{0I}, \tag{3.19}$$

where $\bar{u}_{0I}, \bar{v}_{0I}, \bar{h}_{0I}$ satisfy (3.17). It may be readily checked that the PV $\tilde{\zeta}_0 - \tilde{h}_0$ of the fast component is identically zero.

The fast component satisfying the KG equation with well-posed initial conditions describes the inertia–gravity waves due to the unbalanced $(\tilde{u}_{0I}, \tilde{v}_{0I}, \tilde{h}_{0I})$ part of the initial disturbance and propagating out of it. The corresponding solution may be written as follows:

$$\tilde{h}_0(\boldsymbol{x};t) = \sum_{\pm} \int \mathrm{d}\boldsymbol{k}\, H_0^{(\pm)}(\boldsymbol{k}) \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} \pm \Omega_k t)} + \tilde{h}_{00}, \tag{3.20}$$

where

$$H_0^{(\pm)}(\boldsymbol{k}) = \frac{1}{2}\left(\hat{\tilde{h}}_{0I}(\boldsymbol{k}) \pm \mathrm{i}\frac{\hat{D}_I(\boldsymbol{k})}{\Omega_{\boldsymbol{k}}}\right) \tag{3.21}$$

and

$$\tilde{h}_{00} = \sum_{\pm} \int \mathrm{d}\boldsymbol{k}\, H_{00}^{(\pm)}(\boldsymbol{k}, t_1, t_2, \ldots) \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} \pm \Omega_k t)}. \tag{3.22}$$

Here and below we use the hat notation for Fourier-transforms. The function $H_{00}^{(\pm)}(\boldsymbol{k}, t_1, t_2, \ldots)$ is as yet unknown. It represents a possible slow-time dependence of the Fourier-amplitude of the wave solution and the only condition to be satisfied by $H_{00}^{(\pm)}$ is that it is zero at $t_1 = t_2 = \ldots 0$. One can readily verify that the full time-derivative of $\tilde{h}_{00}$ is expressed as $\partial_t \tilde{h}_{00}|_{t=0} = \epsilon \partial_{t_1} \tilde{h}_{00}|_{t=0} + O(\epsilon^2)$ and, therefore, $\tilde{h}_{00}$ makes no contribution to the lowest-order initial conditions (3.14), (3.15). This additional term in the zeroth-order wave solution for the free-surface elevation and the analogous term in the velocity solution below are, in principle, necessary to avoid resonances at higher orders of the perturbation theory.

The solution for the velocity field is most conveniently written with the help of the complex notation $\mathscr{U} = u + \mathrm{i}v$ which will be also extensively used below in §4. The KG equation

$$-\frac{\partial^2 \tilde{\mathscr{U}}_0}{\partial t^2} - \tilde{\mathscr{U}}_0 + \nabla^2 \tilde{\mathscr{U}}_0 = 0 \tag{3.23}$$

with the initial conditions

$$\tilde{\mathscr{U}}_0|_{t=0} = \tilde{u}_{0I} + \mathrm{i}\tilde{v}_{0I} \equiv \tilde{\mathscr{U}}_{0I}, \tag{3.24}$$

$$\partial_t \tilde{\mathscr{U}}_0|_{t=0} \equiv \mathscr{W}_I = -\mathrm{i}\tilde{\mathscr{U}}_{0I} - \left(\partial_x \tilde{h}_{0I} + \mathrm{i}\partial_y \tilde{h}_{0I}\right) \tag{3.25}$$

results and is solved by

$$\tilde{\mathscr{U}}_0(\boldsymbol{x};t) = \sum_{\pm} \int \mathrm{d}\boldsymbol{k}\, U_0^{(\pm)}(\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} \pm \Omega_k t)} + \tilde{\mathscr{U}}_{00} \tag{3.26}$$

with

$$U_0^{(\pm)}(\boldsymbol{k}) = \frac{1}{2}\left(\hat{\tilde{\mathscr{U}}}_{0I}(\boldsymbol{k}) \pm \mathrm{i}\,\frac{\hat{\mathscr{W}}_I(\boldsymbol{k})}{\Omega_{\boldsymbol{k}}}\right) \tag{3.27}$$

and a slow time-dependent contribution $\tilde{\mathscr{U}}_{00}$ analogous to $\tilde{h}_{00}$.

Equations (3.20), (3.26) represent outgoing dispersive waves and, hence, at any fixed point $\boldsymbol{x} = (x, y)$

$$\tilde{h}_0(\boldsymbol{x};t) = O\left(\frac{1}{t}\right), \quad \tilde{\mathscr{U}}_0(\boldsymbol{x};t) = O\left(\frac{1}{t}\right), \quad t \to \infty \tag{3.28}$$

(see Appendix A for the proof). This, in turn, means that the fast-time averages of the fast components at a given point on the plane (x, y) vanish:

$$\langle \tilde{u}_0(x,y)\rangle = \langle \tilde{v}_0(x,y)\rangle = \langle \tilde{h}_0(x,y)\rangle = 0, \tag{3.29}$$

where for any function $f(t)$ the average is defined as

$$\langle f\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t.$$

Thus, at the zeroth order of the perturbation theory we get a fast–slow motion splitting defined in a unique way starting from arbitrary initial conditions. Note that the procedure imposes no *à priori* limitations on the relative initial values of the fast and the slow components. The fast part of the flow is completely resolved while the slow part remains undetermined. Its evolution equation comes from the condition of the absence of secular growth of the next-order solution.

3.2. *The second order of the perturbation theory*

Equations (2.18) and (2.20) give at this order

$$\partial_t \boldsymbol{v}_1 + \hat{\boldsymbol{z}} \wedge \boldsymbol{v}_1 = -\nabla h_1 + \mathscr{R}_{\boldsymbol{v}}^{(0)}, \qquad \mathscr{R}_{\boldsymbol{v}}^{(0)} = \left(\mathscr{R}_u^{(0)}, \mathscr{R}_v^{(0)}\right) = -\left(\partial_{t_1} + \boldsymbol{v}_0\cdot\nabla\right)\boldsymbol{v}_0. \tag{3.30}$$

The first-order PV equation is

$$\partial_t\,(\zeta_1 - h_1) - \Pi_0\,\partial_t\tilde{h}_0 + \tilde{u}_0\partial_x\Pi_0 + \tilde{v}_0\partial_y\Pi_0 = -\partial_{t_1}\Pi_0 - J(\bar{h}_0, \Pi_0). \tag{3.31}$$

A consistency condition for having solutions of (3.31) bounded in time is obtained by applying the fast-time averaging to the both sides of (3.31) and gives the standard Charney–Obukhov quasi-geostrophic PV equation (QGPV) (cf. (2.15))

$$\partial_{t_1}\Pi_0 + J(\bar{h}_0, \Pi_0) \equiv \partial_{t_1}(\nabla^2\bar{h}_0 - \bar{h}_0) + J(\bar{h}_0, \nabla^2\bar{h}_0) = 0. \tag{3.32}$$

This equation, together with the initial condition (3.13) and the geostrophic relations (3.17) defines completely the evolution of the slow component of motion on the time scale t_1 which is a typical advection time of localized initial disturbances.

Using (3.32) and (3.18) we can integrate (3.31) once in t and get

$$\zeta_1 - h_1 = \Pi_1(x, y; t_1, \ldots) + \mathscr{R}_\zeta^{(0)}, \quad \mathscr{R}_\zeta^{(0)} = \Pi_0\tilde{h}_0 - \tilde{u}_0\partial_y\Pi_0 + \tilde{v}_0\partial_x\Pi_0 - J(\tilde{\mathscr{G}}_0, \Pi_0), \tag{3.33}$$

with

$$\langle \mathscr{R}_\zeta^{(0)}\rangle = 0. \tag{3.34}$$

Here $\Pi_1(x, y; t_1, \ldots)$ is an as yet undetermined slow function and we have introduced the primitive of $\tilde{h}_0$ and its variation with respect to the mean:

$$\tilde{\mathscr{H}}_0 = \int_0^t \tilde{h}_0(t')\,\mathrm{d}t', \qquad \tilde{\mathscr{G}}_0 \equiv \tilde{\mathscr{H}}_0 - \langle\tilde{\mathscr{H}}_0\rangle. \tag{3.35}$$

The function $\tilde{\mathscr{H}}_0$ satisfies the inhomogeneous KG equation

$$-\frac{\partial^2\tilde{\mathscr{H}}_0}{\partial t^2} - \tilde{\mathscr{H}}_0 + \nabla^2\tilde{\mathscr{H}}_0 = -\partial_t\tilde{h}_0|_{t=0} = D_I \tag{3.36}$$

with the initial conditions

$$\tilde{\mathscr{H}}_0|_{t=0} = 0, \qquad \partial_t\tilde{\mathscr{H}}_0|_{t=0} = \tilde{h}_{0I}. \tag{3.37}$$

For the time average of $\tilde{\mathscr{H}}_0$ we have

$$-\langle\tilde{\mathscr{H}}_0\rangle + \nabla^2\langle\tilde{\mathscr{H}}_0\rangle = D_I \tag{3.38}$$

and, hence, $\langle \tilde{\mathscr{H}}_0 \rangle \neq 0$ if $D_I \neq 0$. Correspondingly, $\tilde{\mathscr{G}}_0$ satisfies the following initial-value problem:

$$-\frac{\partial^2 \tilde{\mathscr{G}}_0}{\partial t^2} - \tilde{\mathscr{G}}_0 + \nabla^2 \tilde{\mathscr{G}}_0 = 0, \qquad \tilde{\mathscr{G}}_0|_{t=0} = -\langle \tilde{\mathscr{H}}_0 \rangle, \qquad \partial_t \tilde{\mathscr{G}}_0|_{t=0} = \tilde{h}_{0I}. \tag{3.39}$$

Thus, the first-order analog of (3.5), (3.6), (3.7) consists of (3.33) and (cf. (3.30))

$$\partial_t \zeta_1 + D_1 = \partial_x \mathscr{R}_v^{(0)} - \partial_y \mathscr{R}_u^{(0)} \equiv \mathscr{Z}, \tag{3.40}$$

$$\partial_t D_1 - \zeta_1 = -\nabla^2 h_1 + \partial_x \mathscr{R}_u^{(0)} + \partial_y \mathscr{R}_v^{(0)} \equiv -\nabla^2 h_1 + \mathscr{D}. \tag{3.41}$$

The explicit expressions for $\mathscr{Z}, \mathscr{D}$ are

$$\mathscr{Z} = -[\partial_{t_1} \zeta_0 + \boldsymbol{v}_0 \cdot \nabla \zeta_0 + \zeta_0 D_0], \tag{3.42}$$

$$\mathscr{D} = -[\partial_{t_1} D_0 + \boldsymbol{v}_0 \cdot \nabla D_0 + (\partial_x u_0)^2 + (\partial_y v_0)^2 + 2\partial_x v_0\, \partial_y u_0]. \tag{3.43}$$

From (3.33), (3.40), (3.41) we get the following equation for the first correction to the free-surface elevation:

$$-\frac{\partial^2 h_1}{\partial t^2} - h_1 + \nabla^2 h_1 = \frac{\partial^2 \mathscr{R}_\zeta^{(0)}}{\partial t^2} + \mathscr{R}_\zeta^{(0)} + \mathscr{D} - \partial_t \mathscr{Z} + \Pi_1. \tag{3.44}$$

Now one can proceed as before by splitting h_1 into a sum of slow and fast components

$$h_1 = \tilde{h}_1(x, y; t, \ldots) + \bar{h}_1(x, y; t_1, \ldots) \tag{3.45}$$

and averaging (3.44) with respect to t. The following equations for $\tilde{h}_1, \bar{h}_1$ result:

$$-\frac{\partial^2 \tilde{h}_1}{\partial t^2} - \tilde{h}_1 + \nabla^2 \tilde{h}_1 = \frac{\partial^2 \mathscr{R}_\zeta^{(0)}}{\partial t^2} + \mathscr{R}_\zeta^{(0)} + \mathscr{D} - \langle \mathscr{D} \rangle - \partial_t \mathscr{Z}. \tag{3.46}$$

$$-\bar{h}_1 + \nabla^2 \bar{h}_1 = \Pi_1 + \langle \mathscr{D} \rangle, \tag{3.47}$$

where from the definition of $\mathscr{D}$ we obtain

$$\langle \mathscr{D} \rangle = 2[\partial_x^2 \bar{h} \partial_y^2 \bar{h} - (\partial_{xy}^2 \bar{h})^2]. \tag{3.48}$$

The initial conditions for the full first-order fields are null, as follows from the fact that the initial conditions for velocity and free-surface elevation are already satisfied by zero-order fields (cf. (3.4)). From this and using the definition of Π_1 (3.33) we obtain

$$\Pi_1|_{t=0} = -\mathscr{R}_\zeta^{(0)}|_{t=0} = -[\Pi_I \tilde{h}_{0I} - \tilde{u}_{0I} \partial_y \Pi_I + \tilde{v}_{0I} \partial_x \Pi_I + J(\langle \tilde{\mathscr{H}}_0 \rangle, \Pi_I)]. \tag{3.49}$$

The initial value $\bar{h}_{1I} = \bar{h}_1|_{t=0}$ can be determined from

$$-\bar{h}_{1I} + \nabla^2 \bar{h}_{1I} = \langle \mathscr{D} \rangle|_{t_1=0} + \Pi_1|_{t_1=0}, \tag{3.50}$$

where the right-hand side of this equation is a function of initial values u_I, v_I, h_I only, as follows from (3.38), (3.48), (3.49). To solve the inhomogeneous KG equation (3.46) we need initial conditions for $\tilde{h}_1$ and its time-derivative. The first one is, obviously,

$$\tilde{h}_1|_{t=0} \equiv \tilde{h}_{1I} = -\bar{h}_{1I}. \tag{3.51}$$

The second one follows from (3.33), (3.40) using the fact that $D_1|_{t=0} = 0$:

$$\partial_t \tilde{h}_1|_{t=0} = \mathscr{Z}|_{t=0} - \partial_t \mathscr{R}_\zeta^{(0)}|_{t=0} \tag{3.52}$$

with the right-hand side entirely expressed in terms of initial fields using, where necessary, the evolution equation for the slow component. We, thus, have an inhomogeneous linear initial-value problem for $\tilde{h}_1$ with the source term which may be schematically rewritten as

$$-\frac{\partial^2 \tilde{h}_1}{\partial t^2} - \tilde{h}_1 + \nabla^2 \tilde{h}_1 = -2\partial_{t_1} D_0 + \mathscr{F}^{(s)}(\boldsymbol{x})\mathscr{F}_0^{(f)}(\boldsymbol{x};t) + \mathscr{F}_1^{(f)}(\boldsymbol{x};t)\mathscr{F}_2^{(f)}(\boldsymbol{x};t), \tag{3.53}$$

where the superscripts denote slow and fast spatially localized functions, with $\mathscr{F}_i^{(f)}(\boldsymbol{x};t)$, $i = 0,1,2$, being some solutions of the homogeneous KG equation – see (3.5), (3.33), (3.42), (3.43), (3.46), (3.48) for their precise expressions.

The initial problem (3.53), (3.51), (3.52) can be straightforwardly solved. It is shown in Appendix A that the second and the third terms on the right-hand side of (3.53) are not resonant, i.e. it is non-divergent when integrated in space and time with the inverse KG operator. The term with the slow-time derivative of the divergence is resonant if the functions $\tilde{h}_{00}, \tilde{\mathscr{U}}_{00}$ in (3.20), (3.26) depend on the first slow time t_1. Obviously, in order to avoid resonance it is sufficient to suppose that they do not. Thus, the initial-value problem is well-posed and requires no additional constraints, i.e. there is no slow modulation of the wave envelope at this order. The solution represents the inertia–gravity waves generated by the source term in (3.46) and decays at least as $O(1/t)$ as $t \to \infty$ at any given point on the plane.

Thus, the first order of the perturbation theory provides the evolution equation for the zeroth-order slow field and a (still fast) correction to the zeroth-order fast (wave) field. The evolution of the first correction to the slow field is not determined at this order. It is, however, important to note that it is influenced by the unbalanced part of the flow via initial conditions, cf. (3.50).

3.3. *The third order of the perturbation theory*

At this order the second slow time t_2 enters and the horizontal momentum equations read

$$\partial_t \boldsymbol{v}_2 + \hat{\boldsymbol{z}} \wedge \boldsymbol{v}_2 = -\nabla h_2 + \mathscr{R}_{\boldsymbol{v}}^{(1)}, \qquad \mathscr{R}_{\boldsymbol{v}}^{(1)} = -(\partial_{t_2} + \boldsymbol{v}_1 \cdot \nabla)\boldsymbol{v}_0 - (\partial_{t_1} + \boldsymbol{v}_0 \cdot \nabla)\boldsymbol{v}_1. \tag{3.54}$$

The PV equation gives

$$\partial_t q_2 + \partial_{t_1} q_1 + \partial_{t_2} q_0 + \boldsymbol{v}_0 \cdot \nabla q_1 + \boldsymbol{v}_1 \cdot \nabla q_0 = 0, \tag{3.55}$$

where

$$q_0 = \zeta_0 - h_0 = \Pi_0, \tag{3.56}$$

$$q_1 = \zeta_1 - h_1 - h_0\Pi_0 = \Pi_1 - h_0\Pi_0 + \mathscr{R}_\zeta^{(0)}, \tag{3.57}$$

$$q_2 = \zeta_2 - h_2 - h_0(\zeta_1 - h_1) + (h_0^2 - h_1)\Pi_0. \tag{3.58}$$

The main goal of this subsection is to obtain corrections to the slow-component evolution equation (3.32). We will not, therefore, pursue the calculation of the wave-field corrections at this order, which is a matter of cumbersome but straightforward calculations because, again, nonlinear (i.e. due to the terms $\boldsymbol{v}_1\nabla\boldsymbol{v}_0 + \boldsymbol{v}_0\nabla\boldsymbol{v}_1$ on the right-hand side of (3.54)) resonances are absent, cf. Appendix A. As for the linear resonances due to the terms $\partial_{t_2}\boldsymbol{v}_0 + \partial_{t_1}\boldsymbol{v}_1$, they pose no problems and are eliminated with the help of the slow-time dependent additions $\tilde{h}_{00}$ and $\tilde{\mathscr{U}}_{00}$ to the free-wave field and analogous contributions which may be added to $\tilde{h}_1$, $\tilde{u}_1$. We, thus, concentrate on the PV equation (3.55). Let us average this equation in t and consider it term by term. We have

$$\langle \partial_t q_2 \rangle = 0. \tag{3.59}$$

By virtue of (3.29), (3.34)

$$\langle \partial_{t_1} q_1 \rangle = \partial_{t_1}(\Pi_1 - \bar{h}_0 \Pi_0). \tag{3.60}$$

By splitting velocity $\boldsymbol{v}_0$ and q_1 into fast and slow components which are denoted, as usual, by a tilde and overbar respectively, we get

$$\langle \boldsymbol{v}_0 \cdot \nabla q_1 \rangle = \bar{\boldsymbol{v}}_0 \cdot \nabla \bar{q}_1 + \langle \tilde{\boldsymbol{v}}_0 \cdot \nabla \tilde{q}_1 \rangle. \tag{3.61}$$

The fast-time averaging of expressions containing rapidly oscillating wave factors may give a non-zero (or divergent) result only if the expressions in question decay slow enough at $t \to \infty$. As follows from (3.28) and (3.33), for any fixed point of the plane in this limit

$$\mathscr{R}_{\zeta}^{(0)} = O\left(\frac{1}{t}\right) \Rightarrow \tilde{q}_1 = O\left(\frac{1}{t}\right) \tag{3.62}$$

and hence

$$\tilde{\boldsymbol{v}}_0 \cdot \nabla \tilde{q}_1 = O\left(\frac{1}{t^2}\right) \Rightarrow \langle \tilde{\boldsymbol{v}}_0 \cdot \nabla \tilde{q}_1 \rangle = 0. \tag{3.63}$$

Finally,

$$\langle \boldsymbol{v}_1 \cdot \nabla q_0 \rangle = \langle \boldsymbol{v}_1 \rangle \cdot \nabla \Pi_0. \tag{3.64}$$

It readily follows from (3.46) that

$$-\langle \tilde{h}_1 \rangle + \nabla^2 \langle \tilde{h}_1 \rangle = 0 \tag{3.65}$$

and, hence, $\langle \tilde{h}_1 \rangle = 0$ with our choice of decaying boundary conditions on the plane. We thus get from (3.28), (3.30), (3.65)

$$\langle \boldsymbol{v}_1 \rangle = \bar{\boldsymbol{v}}_1 = \hat{\boldsymbol{z}} \wedge [\nabla \bar{h}_1 + (\partial_{t_1} + \bar{\boldsymbol{v}}_0 \cdot \nabla)\bar{\boldsymbol{v}}_0]. \tag{3.66}$$

Using (3.59)–(3.66) we arrive at the following slow equation:

$$(\partial_{t_2} + \bar{\boldsymbol{v}}_1 \cdot \nabla)\Pi_0 + (\partial_{t_1} + \bar{\boldsymbol{v}}_0 \cdot \nabla)(\Pi_1 - \bar{h}_0 \Pi_0) = 0. \tag{3.67}$$

So, remarkably, there are no fast–fast (wave-drag) contributions to (3.67) which thus may be safely obtained by a direct expansion of (2.20) in ϵ considering all variables as being slow. In order to get a closed equation for $\bar{h}_0, \bar{h}_1$ we express Π_1 as

$$\Pi_1 = \partial_x \bar{v}_1 - \partial_x \bar{u}_1 - \bar{h}_1, \tag{3.68}$$

according to its definition (3.33), (3.34). By using (3.66) which, with the help of the zeroth-order geostrophic balance conditions (3.17), may be rewritten as

$$\bar{\boldsymbol{v}}_1 = \hat{\boldsymbol{z}} \wedge \nabla \bar{h}_1 - \partial_{t_1} \nabla \bar{h}_0 - J(\bar{h}_0, \nabla \bar{h}_0) \tag{3.69}$$

we get

$$\Pi_1 = \nabla^2 \bar{h}_1 - \bar{h}_1 - 2J(\partial_x \bar{h}_0, \partial_y \bar{h}_0). \tag{3.70}$$

With the help of (3.69) and the evolution equation for Π_0, equation (3.67) takes the following form:

$$\begin{aligned} &\partial_{t_2}\Pi_0 + \partial_{t_1}(\nabla^2 \bar{h}_1 - \bar{h}_1 - 2J(\partial_x \bar{h}_0, \partial_y \bar{h}_0) - \bar{h}_0 \Pi_0 - \nabla \bar{h}_0 \cdot \nabla \Pi_0) \\ &\quad + J(\bar{h}_0, \nabla^2 \bar{h}_1 - \bar{h}_1 - 2J(\partial_x \bar{h}_0, \partial_y \bar{h}_0) - \bar{h}_0 \Pi_0 - \nabla \bar{h}_0 \cdot \nabla \Pi_0) \\ &\quad + J\left(\bar{h}_1 - \frac{(\nabla \bar{h}_0)^2}{2}, \Pi_0\right) = 0. \end{aligned} \tag{3.71}$$

This equation describes the next-order correction to the equation (3.32) which is necessary to take into account when studying the slow evolution of the balanced component of the flow for times much longer than t_1. One can combine the two equations by introducing a 'full' slow elevation $\bar{h} = \bar{h}_0 + \epsilon\bar{h}_1$ and a 'full' slow-time derivative $\partial_\tau = \partial_{t_1} + \epsilon\partial_{t_2}$. The result, up to an $O(\epsilon^2)$ correction, is the following 'improved' QGPV equation:

$$\frac{\mathrm{D}}{\mathrm{D}\tau}[\nabla^2\bar{h} - \bar{h} - \epsilon\bar{h}(\nabla^2\bar{h} - \bar{h}) - \epsilon\nabla\bar{h}\cdot\nabla(\nabla^2\bar{h} - \bar{h}) - 2\epsilon J(\partial_x\bar{h}, \partial_y\bar{h})] = 0, \tag{3.72}$$

where

$$\frac{\mathrm{D}}{\mathrm{D}\tau}(\dots) := \partial_\tau(\dots) + J\left(\bar{h} - \epsilon\frac{(\nabla\bar{h})^2}{2}, \dots\right). \tag{3.73}$$

This equation is of the 'generalized vorticity' kind. It should be solved with the initial conditions for the full height following from (3.13), (3.50) and may be rewritten in alternative form(s) by using the lower-order QGPV equation. It coincides with the balanced dynamics equation obtained by imposing 'height-slaving' (Warn *et al.* 1995, equation (63)) which, in turn, reproduces the iterated geostrophic model IG2 derived earlier by Allen (1993). Note, however, that no slaving or other constraint was imposed in the present derivation which means that we prove the validity of this model assuming only the smallness of the Rossby number and single-scaledness of the initial perturbation.

3.4. *Discussion of the QG results on the f-plane*

Thus, at the leading orders in ϵ the motion in the QG regime is split in a unique and operational way into slow and fast components. The former corresponds to the *linearized* evolution of PV made up of the *full* initial data according to the QGPV equation. The latter represents inertia–gravity waves rapidly leaving the initial disturbance. The splitting is complete in the sense that there is no interaction at all between the fast and the slow components, which feel each other's presence only through initial conditions. In fact, at the lowest order this result may be anticipated from the direct calculation of the energy of the slow–fast interaction. Indeed, the energy of the RSW motion is given by

$$E = \int \mathrm{d}x\,\mathrm{d}y\left(H\frac{\boldsymbol{v}^2}{2} + g\frac{H^2}{2}\right) \tag{3.74}$$

and under the hypotheses of the QG scaling and zero total mass anomaly becomes

$$E = \int \mathrm{d}x\,\mathrm{d}y\left[(1+\epsilon h)\frac{\boldsymbol{v}^2}{2} + \frac{h^2}{2}\right]. \tag{3.75}$$

Using the slow–fast splitting (3.9), (3.16) one gets from (3.75) at the lowest order in ϵ

$$E = \bar{E} + \tilde{E} + E_{int}, \tag{3.76}$$

where

$$E_{int} = -\frac{1}{2}\int \mathrm{d}x\,\mathrm{d}y\,\bar{h}_0\,(\tilde{\zeta} - \tilde{h}_0) \tag{3.77}$$

and $\bar{E}$ and $\tilde{E}$ denote pure slow-component and fast-component energies, respectively. The fast–slow interaction energy density thus vanishes, being proportional to the potential vorticity of the fast component which is identically zero. What is more surprising, the splitting persists at the next orders, as we have shown. Although at the present time we are not able to prove this at all orders, this hypothesis seems plausible even at high orders of the perturbation theory due to the presence of the spectral gap in the RSW model – see Appendix A.

Note the difference between our proof and that of Embid & Majda (1996) and Babin *et al.* (1998*a, b*): in our case of an open domain and broad-spectrum disturbances no wave–wave and wave–mean flow resonances occur at low orders. This explains the absence in our analysis of the quadruplet wave interactions found by Babin *et al.* (1998*a, b*). The physical reason for this is clear: the fast waves are propagating out of the disturbance, and do not stay at a given space point long enough to produce a resonance, which is not true in a periodic box.

In the context of the slow-manifold studies our approach is close to that of Medvedev (1999) where the slow-manifold was found perturbatively (with geostrophic scaling) in nonlinearity. However, by construction we do not impose any special 'slaving' – the fast and the slow components coexist without influencing each other while the former is being radiating away which, again, shows the difference between the present approach and the normal-mode one.

4. Nonlinear geostrophic adjustment in the FD regime

4.1. *The lowest order of the perturbation theory*

Let us recall that in the FD regime the surface elevation variable h represents the *full* depth of the fluid and, thus, is always greater than zero. At the lowest order we have from (2.18), (2.19)

$$\partial_t \boldsymbol{v}_0 + \hat{\boldsymbol{z}} \wedge \boldsymbol{v}_0 = -\nabla h_0, \tag{4.1}$$

$$\partial_t h_0 = 0, \tag{4.2}$$

with the same initial conditions (3.4) as before. It follows from (4.2) that $h_0 \equiv \bar{h}_0$ has no fast component in this regime (which is natural because at the lowest order in ϵ the PV in this regime is a function of $\bar{h}_0$ only – cf. (2.20)). Solutions to (4.1), (4.2) may be immediately split into fast and slow components, denoted as usual by an overbar and tilde with the slow one satisfying the geostrophic balance (3.17) and the fast one satisfying the homogeneous version of (4.1). The initial conditions are split as

$$\bar{h}_0|_{t=0} = h_I, \qquad \bar{\boldsymbol{v}}|_{t=0} = \boldsymbol{v}_I^{(g)}, \qquad \tilde{\boldsymbol{v}}|_{t=0} = \boldsymbol{v}_I^{(a)}, \tag{4.3}$$

where we separated the geostrophic $\boldsymbol{v}_I^{(g)}$ and the ageostrophic $\boldsymbol{v}_I^{(a)}$ part of the initial conditions with respect to the initial free-surface elevation:

$$\boldsymbol{v}_I^{(g)} = \hat{\boldsymbol{z}} \wedge \nabla h_I, \qquad \boldsymbol{v}_I^{(a)} = \boldsymbol{v}_I - \boldsymbol{v}_I^{(g)}. \tag{4.4}$$

The homogeneous version of (4.1) describes non-propagating inertial oscillations corresponding to the small-$\boldsymbol{k}$ limit of the dispersion equation (2.10). Using the complex notation $\mathscr{U}$, which is especially convenient in this case (cf. e.g. Young & Ben Jelloul 1997 and see below), the corresponding solution for the fast velocity field is

$$\tilde{\mathscr{U}}(\boldsymbol{x};t) = \mathscr{A}_0(\boldsymbol{x};t_1,\ldots)\mathrm{e}^{-\mathrm{i}t}, \qquad \mathscr{A}_0(\boldsymbol{x};t_1,\ldots)|_{t=0} = \mathscr{U}_I^{(a)}, \tag{4.5}$$

where $\mathscr{A}_0$ is an amplitude (possibly slowly modulated) of the inertial oscillations. Note the following simple relations which we use below:

$$\int \mathrm{d}t\, \tilde{\mathscr{U}}_0(t) = \mathrm{i}\,\tilde{\mathscr{U}}_0(t), \qquad \int \mathrm{d}t\, \tilde{\mathscr{U}}_0^*(t) = -\mathrm{i}\,\tilde{\mathscr{U}}_0^*(t) \tag{4.6}$$

(the star denotes the complex conjugate). The velocity may be rewritten in fully complex notation using the complex variables $z = x + \mathrm{i}y, z^* = x - \mathrm{i}y$ instead of x, y:

$$\mathscr{U}_0 = \tilde{\mathscr{U}}_0 + \bar{\mathscr{U}}_0 = \tilde{\mathscr{U}}_0 + (-\partial_y \bar{h}_0 + \mathrm{i}\partial_x \bar{h}_0) = \mathscr{A}_0 \mathrm{e}^{-\mathrm{i}t} + 2\mathrm{i}\partial_{z^*}\bar{h}_0, \tag{4.7}$$

where it is useful to recall the basic formulas

$$\partial_z = \tfrac{1}{2}(\partial_x - \mathrm{i}\partial_y), \qquad \partial_{z^*} = \tfrac{1}{2}(\partial_x + \mathrm{i}\partial_y), \qquad \nabla^2 = 4\partial^2_{zz^*}, \qquad J(z, z^*) = -2\mathrm{i}. \tag{4.8}$$

The advective derivative and divergence are expressed in complex terms as

$$\boldsymbol{v} \cdot \nabla = \mathscr{U}\partial_z + \mathscr{U}^*\partial_{z^*}, \tag{4.9}$$

$$\nabla \cdot \boldsymbol{v} = \partial_z \mathscr{U} + \partial_{z^*}\mathscr{U}^*. \tag{4.10}$$

Thus, at the leading order of the perturbation theory in the FD regime, the fast and the slow components of motion are split, the former describing inertial oscillations resulting from the imbalanced part of the initial conditions, while the latter evolving from the geostrophically balanced part of initial conditions remains undetermined.

4.2. *The second order of the perturbation theory*

At this order we have the same equations (3.30) as in § 3.2 for the velocity field. The evolution of the h-field is given by

$$\partial_t h_1 = -\partial_{t_1}\bar{h}_0 - \nabla \cdot (\bar{h}_0 \boldsymbol{v}_0^{(a)}) \tag{4.11}$$

because the contribution of the balanced velocity on the right-hand side vanishes identically. The t-independent first term on the right-hand side is trivially resonant as it leads to a linear growth in time of h_1 and, hence, should be equal to zero which means that the characteristic time scale of h_0 is, in fact, t_2 (this result is consistent with scaling leading to (2.17) – cf. Stegner & Zeitlin 1995). Having eliminated the resonance, equation (4.11) can be immediately integrated to give, in complex notation,

$$h_1 = \tilde{h}_1 + \bar{h}_1 = -\mathrm{i}\partial_z(\bar{h}_0\tilde{\mathscr{U}}_0) + \mathrm{i}\partial_{z^*}(\bar{h}_0\tilde{\mathscr{U}}_0^*) + \bar{h}_1, \tag{4.12}$$

where we used (4.6). Thus, we get the following equation for $\mathscr{U}_1$:

$$\partial_t \mathscr{U}_1 + \mathrm{i}\,\mathscr{U}_1 = -2\mathrm{i}\partial_{z^*}[-\partial_z(\bar{h}_0\tilde{\mathscr{U}}_0) + \partial_{z^*}(\bar{h}_0\tilde{\mathscr{U}}_0^*) - \mathrm{i}\,\bar{h}_1] \\ -(\partial_{t_1}\mathscr{U}_0 + \mathscr{U}_0\partial_z\mathscr{U}_0 + \mathscr{U}_0^*\partial_{z^*}\mathscr{U}_0). \tag{4.13}$$

This is an inhomogeneous equation for inertial oscillations which are non-propagative and non-dispersive, unlike the inertia–gravity waves appearing at the same stage in the QG regime above. Hence, if forced at its proper frequency, the equation exhibits a typical resonance behaviour. The resonant forcing $\sim \mathrm{e}^{-\mathrm{i}t}$ corresponds to terms containing $\mathscr{U}_0$ only once on the right-hand side of (4.13). These terms must be eliminated in order to avoid a secular growth of $\mathscr{U}_1$. Thus, we obtain the following modulation equation for the amplitude of inertial oscillations:

$$\partial_{t_1}\mathscr{A}_0 - 2\mathrm{i}\partial^2_{zz^*}(\bar{h}_0\mathscr{A}_0) + 2\mathrm{i}\mathscr{A}_0\partial^2_{zz^*}\bar{h}_0 + 2\mathrm{i}(\partial_{z^*}\bar{h}_0\partial_z\mathscr{A}_0 - \partial_z\bar{h}_0\partial_{z^*}\mathscr{A}_0) = 0, \tag{4.14}$$

or, coming back to the real notation

$$\partial_{t_1}\mathscr{A}_0 + J(\bar{h}_0, \mathscr{A}_0) - \tfrac{1}{2}\mathrm{i}[\nabla^2(\bar{h}_0\mathscr{A}_0) - \mathscr{A}_0\nabla^2\bar{h}_0] = 0. \tag{4.15}$$

In the present context this equation (up to a change of variables) was first proposed by Fal'kovich (1992) and correctly derived by Fal'kovich, Kuznetsov & Medvedev (1994) – see § 4.4 below.

Having eliminated the secular growth, the regular solution of (4.13) satisfying zero initial conditions may be easily found in the form $\mathscr{U}_1 = \bar{\mathscr{U}}_1 + \tilde{\mathscr{U}}_1$, where

$$\bar{\mathscr{U}}_1 = \mathrm{i}(2\partial_{z^*}\bar{h}_1 + J(\bar{h}_0, \bar{\mathscr{U}}_0) + \mathscr{A}_0^*\partial_{z^*}\mathscr{A}_0), \tag{4.16}$$

$$\tilde{\mathscr{U}}_1 = \tilde{\mathscr{U}}_1^+ + \tilde{\mathscr{U}}_1^{--} + \tilde{\mathscr{U}}_1^- \equiv -\tfrac{1}{2}\mathrm{i}\mathscr{C}_1\mathrm{e}^{it} + \mathrm{i}\mathscr{C}_2\mathrm{e}^{-2\mathrm{i}t} + \mathscr{A}_1(t_1, \ldots)e^{-\mathrm{i}t} \tag{4.17}$$

with obvious notation, and, respectively,

$$\left.\begin{aligned} \mathscr{C}_1 &= \partial_{xy}^2(\bar{h}_0\mathscr{A}_0^*) + \mathscr{A}_0^*\partial_{xy}^2\bar{h}_0 + \tfrac{1}{2}\mathrm{i}[\hat{P}(\bar{h}_0\mathscr{A}_0^*) + \mathscr{A}_0^*\hat{P}(\bar{h}_0)], \\ \mathscr{C}_2 &= -\tfrac{1}{2}\mathscr{A}_0(\partial_x\mathscr{A}_0 - \mathrm{i}\partial_y\mathscr{A}_0), \end{aligned}\right\} \tag{4.18}$$

with the operator $\hat{P}$ defined by $\hat{P}(f) := \partial_y^2 f - \partial_x^2 f$.

Thus, at the second order of the perturbation theory the slow component of the motion remains undetermined, being, in fact, even slower than was initially supposed, while the fast component – inertial oscillations – acquires a slow modulation described by a linear Schrödinger-type equation (4.15) with coefficients depending on the slow motion. An induced correction to the slow motion also appears and is given by (4.16).

4.3. *The third order of the perturbation theory*

The horizontal momentum equations are (cf. (3.54))

$$\partial_t \boldsymbol{v}_2 + (\partial_{t_2} + \boldsymbol{v}_1 \cdot \nabla)\boldsymbol{v}_0 + (\partial_{t_1} + \boldsymbol{v}_0 \cdot \nabla)\boldsymbol{v}_1 + \hat{\boldsymbol{z}} \wedge \boldsymbol{v}_2 = -\nabla h_2 \tag{4.19}$$

and to close the system we add the height evolution equation

$$\partial_t h_2 + \partial_{t_1} h_1 + \partial_{t_2}\bar{h}_0 + \nabla(\bar{h}_0\boldsymbol{v}_1 + h_1\boldsymbol{v}_0) = 0. \tag{4.20}$$

Let us average this equation in t supposing boundness of h_2 in time. We obtain

$$\partial_{t_1}\bar{h}_1 + \partial_{t_2}\bar{h}_0 + \nabla\cdot(\bar{h}_0\bar{\boldsymbol{v}}_1 + \bar{h}_1\bar{\boldsymbol{v}}_0) + \langle\nabla(\tilde{h}_1\tilde{\boldsymbol{v}}_0)\rangle = 0. \tag{4.21}$$

Calculating consecutively the bilinear combinations of height and velocity variables in this equation we obtain

$$\begin{aligned} \nabla\cdot(\bar{h}_0\bar{\boldsymbol{v}}_1) = J(\bar{h}_1, \bar{h}_0) - J(\bar{h}_0, \bar{h}_0\nabla^2\bar{h}_0 + \tfrac{1}{2}(\nabla\bar{h}_0)^2) - \tfrac{1}{4}[J(\bar{h}_0, |\tilde{\mathscr{A}}_0|^2 \\ -\mathrm{i}[\nabla\bar{h}_0\cdot(\mathscr{A}_0^*\nabla\mathscr{A}_0 - \mathscr{A}_0\nabla\mathscr{A}_0^*) + \bar{h}_0(\mathscr{A}_0^*\nabla^2\mathscr{A}_0 - \mathscr{A}_0\nabla^2\mathscr{A}_0^*)]], \end{aligned} \tag{4.22}$$

where we used (4.16); and

$$\nabla\cdot(\bar{h}_1\bar{\boldsymbol{v}}_0) = J(\bar{h}_0, \bar{h}_1) \tag{4.23}$$

and

$$\begin{aligned} \langle\nabla(\tilde{h}_1\tilde{\boldsymbol{v}}_0)\rangle = \tfrac{1}{4}[J(\bar{h}_0, |\mathscr{A}_0|^2 \\ -\mathrm{i}[\nabla\bar{h}_0\cdot(\mathscr{A}_0^*\nabla\mathscr{A}_0 - \mathscr{A}_0\nabla\mathscr{A}_0^*) + \bar{h}_0(\mathscr{A}_0^*\nabla^2\mathscr{A}_0 - \mathscr{A}_0\nabla^2\mathscr{A}_0^*)]], \end{aligned} \tag{4.24}$$

where we used (4.12), (4.17). Thus, the contributions containing fast–fast terms remarkably cancel leaving a purely slow evolution equation (2.17) for h_0:

$$\partial_{t_2}\bar{h}_0 - J\left(\bar{h}_0, \bar{h}_0\nabla^2\bar{h}_0 + \tfrac{1}{2}(\nabla\bar{h}_0)^2\right) = 0. \tag{4.25}$$

Thus, again as in the QG case, we have no fast motion drag and the slow component evolves by itself. On the other hand, the fast component evolves on the background of the slow one according to the linear modulation equation (4.15) which may produce non-trivial effects – see below. Knowing the rich physics of the *nonlinear* Schrödinger equation it would be exciting to get a cubic-in-$\mathscr{A}_0$ correction to (4.15) at the present order. This, however, turns out to be impossible. Indeed, such a correction could appear from elimination of resonances in the equation for the next order velocity $\mathscr{U}_2$. Comparing with (4.19) written in complex notation we see that the resonant (i.e. $\sim e^{-it}$) cubic terms appear in the following expressions:

$$\tilde{\mathscr{U}}_0^* \partial_{z^*} \tilde{\mathscr{U}}_1^{--} \tag{4.26}$$

and

$$\tilde{\mathscr{U}}_0 \partial_z \bar{\mathscr{U}}_1 + \bar{\mathscr{U}}_1^* \partial_z \tilde{\mathscr{U}}_0 + \bar{\mathscr{U}}_1 \partial_{z^*} \tilde{\mathscr{U}}_0. \tag{4.27}$$

Combining them we get

$$-\mathrm{i}\, \tilde{\mathscr{U}}_0^* \partial_{z^*} (\tilde{\mathscr{U}}_0 \partial_z \tilde{\mathscr{U}}_0) + \mathrm{i}\, \tilde{\mathscr{U}}_0 \partial_z (\tilde{\mathscr{U}}_0^* \partial_{z^*} \tilde{\mathscr{U}}_0) + \mathrm{i}\, \tilde{\mathscr{U}}_0^* \partial_{z^*} \tilde{\mathscr{U}}_0 \partial_z \tilde{\mathscr{U}}_0 - \mathrm{i}\, \tilde{\mathscr{U}}_0 \partial_z \tilde{\mathscr{U}}_0^* \partial_{z^*} \tilde{\mathscr{U}}_0 \equiv 0. \tag{4.28}$$

Therefore, there are no nonlinear corrections to the Schrödinger-type equation (4.15). An evolution equation for the first correction $\mathscr{A}_1$ to the envelope of inertial oscillations may be easily obtained. We do not display it here because it contains nothing particularly interesting. Note that the absence of cubic corrections in the modulation equation means that there are no corresponding resonant quadruplets of quasi-inertial waves in the vicinity of the threshold frequency f.

4.4. *Discussion of the FD results and a résumé of the NLQG calculations on the f-plane*

Thus, the FD regime is similar to the QG one in the sense that splitting is also taking place and the slow component of the flow evolves according to the balanced equation without being influenced by the fast one. However, in this case the behaviour of the fast component is sensible to the background slow component and is governed by the Schrödinger-type equation (4.15). Nevertheless, the energy of the inertial oscillations

$$\tilde{E} = \int \mathrm{d}x\, \mathrm{d}y\, \tfrac{1}{2} |\mathscr{A}_0|^2 \tag{4.29}$$

is conserved, as well as $\int \mathrm{d}x\, \mathrm{d}y F(\bar{h}_0)\, |\mathscr{A}_0|^2$ where F is arbitrary function of $\bar{h}_0$. Separately, the energy of the slow motion

$$\bar{E} = \int \mathrm{d}x\, \mathrm{d}y\, \tfrac{1}{2} \bar{h}_0 (\nabla \bar{h}_0)^2 \tag{4.30}$$

is also conserved which accomplishes the demonstration of splitting.

It is well known that the classical Schrödinger equation describes dispersion of the wave packets. We can make the resemblance between (4.15) and the Schrödinger equation even closer by the change of variables $\mathscr{A}_0 \to \mathscr{B}_0 = \bar{h}_0 \mathscr{A}_0$ which yields

$$\mathrm{i}[\partial_{t_1} \mathscr{B} + J(\bar{h}_0, \mathscr{B})] + \tfrac{1}{2} \bar{h}_0 \nabla^2 \mathscr{B} - \tfrac{1}{2} \nabla^2 \bar{h}_0\, \mathscr{B} = 0. \tag{4.31}$$

We thus obtain a standard Schrödinger equation with a variable 'mass' $\bar{h}_0$, a 'potential' $\sim \nabla^2 \bar{h}_0$ and additional advection by the geostrophic velocity field produced by $\bar{h}_0$. The following qualitative picture of the FD regime is obtained:

Any initial height perturbation evolves very slowly according to the FD dynamics (2.17). A non-geostrophic part of the velocity perturbation is rapidly oscillating with the inertial period and is slowly dispersed according to (4.31) where the coefficients may be considered as being constant in time at the dispersion time scale. It is simultaneously advected by the geostrophic part of velocity. The details of the dispersion process depend on the initial field ('front') $h_I(\boldsymbol{x})$. As in the standard Schrödinger equation case, it is possible, for particular profiles, that some part of the initial wave packet is trapped by the front ('bound states') and evolves together with it. In any case, the presence of the inertial oscillation packet provides a sort of 'fuzzyness' of the front and may have implications for the transport and mixing properties. The details of this analysis and possible applications will be discussed elsewhere.

As already mentioned, the Schrödinger-type equation (4.31) was first derived in Fal'kovich *et al.* (1994). The procedure used in that paper consisted in a perturbative expansion in the Burger number, simultaneous decomposition of any field in the Fourier series in time using the overtones of the inertial period and subsequent truncation (in spite of the title, no Rossby waves were considered). We thus confirm this result but prove, in addition, the absence of the inertial oscillation (called 'inertial-gravity wave condensate' in Fal'kovich *et al.* 1994) drag in the slow-motion equation and so demonstrate splitting at the next order of the perturbation theory. Finally, let us note that propagation of near-inertial oscillations was studied by similar means, but in a different context, by Young & Ben Jelloul (1997).

Let us briefly mention (with technical details given in Appendix B), what happens if one applies the multi-time-scale perturbation theory in the NLQG regime. The procedure closely follows the FD analysis while being technically simpler (note that the basic expansion parameter here is $\lambda, \lambda \sim \epsilon^{1/2}$ and not ϵ, cf. (2.18), (2.19), (2.20). The main result is that the height anomaly evolves according to the NLQG equation (2.16) while the amplitude equation (4.14) at the leading order becomes the standard Schrödinger equation

$$\partial_{t_1}\mathscr{A}_0 - \tfrac{1}{2}\mathrm{i}\nabla^2\mathscr{A}_0 = 0. \qquad (4.32)$$

5. Comments on the influence of the beta-effect

5.1. *QG regime*

Let us introduce the non-dimensional beta-parameter $\bar\beta$ with an explicit small factor in front of it:

$$f = f_0 + \beta y = f_0(1 + \epsilon_\beta \bar\beta y). \qquad (5.1)$$

Here

$$\epsilon_\beta = O\left(\frac{\beta L}{f_0}\right) \ll 1, \qquad \bar\beta = O(1). \qquad (5.2)$$

A third small parameter thus appears in the theory and its value should be fixed with respect to λ and ϵ. For the QG case we choose $\epsilon_\beta \sim \epsilon^2$, i.e. we suppose that nonlinear terms dominate those induced by β in (2.1). This assumption is justified for the intense mesoscale structures (eddies and fronts) in the ocean and atmosphere (see e.g. Reznik & Grimshaw 2001). The PV takes the form (cf. (2.20))

$$q = \frac{\epsilon\zeta - \lambda h + \epsilon^2\bar\beta y}{1 + \lambda h} \qquad (5.3)$$

and by expanding the fields in asymptotic series (3.1) and repeating the procedure of § 3 we arrive to the same results at the two lowest orders of the perturbation theory. The parameter β appears at the third order. The slow-motion equation (3.72) becomes

$$\frac{\mathrm{D}}{\mathrm{D}\tau}[\nabla^2\bar h - \bar h + \epsilon\bar\beta y - \epsilon\bar h(\nabla^2\bar h - \bar h) - \epsilon\nabla\bar h\cdot\nabla(\nabla^2\bar h - \bar h) - 2\epsilon J(\partial_x\bar h, \partial_y\bar h)] = 0 \qquad (5.4)$$

and we see that the β-term and the nonlinear ageostrophic terms give corrections of the same order to the QGPV equation. Therefore, it is (5.4) which is the proper equation to use while studying the impact of the β-effect on the intense mesoscale structures of general form. On the other hand, the standard QGPV equation on the β-plane is suitable only in cases where the initial state possesses a special symmetry, as for example an axisymmetric localized vortex. In this (specific, but practically important) case the β-term is the main factor of vortex evolution while the nonlinear ageostrophic contributions remain smaller for long times (cf. Reznik, Grimshaw & Benilov 2000; Reznik & Grimshaw 2001). We would like to stress here, however, that the nonlinear ageostrophic terms are crucial for understanding such important physical phenomena as e.g. the cyclone–anticyclone asymmetry.

The equation for the fast field $\tilde{h}_2$ is derived along the same lines as (3.53) and has the following form:

$$-\frac{\partial^2 \tilde{h}_2}{\partial t^2} - \tilde{h}_2 + \nabla^2 \tilde{h}_2 = \mathscr{R}_h^{(1)} + \bar{\beta}(\partial^2_{ty}\tilde{h}_0 + \partial_y \tilde{h}_0 - \partial_x \tilde{h}_0) + 2\bar{\beta} y \nabla^2 \tilde{h}_0. \tag{5.5}$$

Here $\mathscr{R}_h^{(1)}$ does not depend on $\bar{\beta}$ and contains non-resonant terms and linear resonant terms analogous to those on the right-hand side of (3.53). These resonances, as before, may be eliminated by adding to $\tilde{h}_0, \tilde{h}_1$ free-wave solutions with slow-time-dependent Fourier amplitudes, cf. (3.20). The same is true for the second group of terms in the right-hand side of (5.5). On the other hand, the last y-dependent term poses a real problem as it is evident that this resonance, being spatially non-uniform, cannot be removed within the single-space-scale asymptotic theory (formally, this term gives the $O(t^2)$ contribution to $\tilde{h}_2$ when $t \to \infty$). Hence, whenever the initial state (2.5) deviates from the geostrophic balance and $\tilde{h}_{0I} \neq 0$ the problem (5.5) is ill-posed on the β-plane. Moreover, if $\tilde{h}_{0I} = 0$, but some initial $\tilde{h}_n$ are not, where $\tilde{h}_n$ is the nth-order fast field, inevitable secular growth arises at order $n + 2$.

Thus, we come to the conclusion that in the case of non-zero β and unbalanced initial conditions the one-space and multiple-time-scale asymptotic procedure adopted in the present study cannot provide a solution of the RSW equations with any given accuracy. However, it still provides a solution within a certain accuracy which is determined by the 'degree of imbalance' of the initial conditions (2.5). Physically, the problem is related to the fact that in times $O(\epsilon^{-2})$ the fast inertia–gravity waves travel a distance $O(\epsilon^{-2})$ and the Coriolis parameter $f_0(1+\epsilon^2\bar{\beta}y)$ cannot be considered constant along this distance, which was implicitely assumed while writing (3.1). A self-consistent asymptotic procedure on the β-plane (yet to be developed) should take into account a distortion of the fast wave rays due to the spatial inhomogeneity produced by the β-effect. Another way to avoid this difficulty is to consider motion in a zonal channel on the β-plane. There are many reasons to believe, however, that a change of the perturbative scheme along these lines would affect only the fast component of the motion, the slow one remaining unchanged. This assertion rests on two observations: (i) the spectral gap persists in the presence of the β-effect, i.e. inertia–gravity waves are still much faster than Rossby waves; (ii) the 'naive' asymptotic expansion used above already provides a self-consistent derivation of the slow motion equations in the case of a balanced initial state.

5.2. *FD regime*

In the FD regime the characteristic scale is much greater than the deformation radius and we, therefore, take the β-parameter to be of the same order of magnitude as the

Rossby number:

$$\epsilon_\beta = O(\epsilon). \tag{5.6}$$

In this case the lowest order of the perturbation theory does not change and the formulas of § 4.1 remain valid. The β-effect manifests itself at the second order and the Schrödinger equation (4.15) acquires a contribution with a y-dependent coefficient:

$$\partial_{t_1}\mathscr{A}_0 + J(\bar{h}_0, \mathscr{A}_0) - \tfrac{1}{2}\mathrm{i}[\nabla^2(\bar{h}_0\mathscr{A}_0) - \mathscr{A}_0\nabla^2\bar{h}_0] + \mathrm{i}\bar{\beta}y\mathscr{A}_0 = 0. \tag{5.7}$$

The slow-time evolution of the free-surface elevation also changes (cf. e.g. Stegner & Zeitlin 1995) and becomes (cf. (4.25))

$$\partial_{t_2}\bar{h}_0 - \bar{\beta}\bar{h}_0\partial_x\bar{h}_0 - J\left(\bar{h}_0, \bar{h}_0\nabla^2\bar{h}_0 + \tfrac{1}{2}(\nabla\bar{h}_0)^2\right) = 0. \tag{5.8}$$

Note, first, that we still find splitting on the β-plane (cf. Fal'kovich *et al.* 1994, where the presence of the fast-component drag in the slow-motion equation was claimed on a heuristic level; this is not confirmed by our calculations). Second, the presence of the 'simple wave' β-induced term in the FD equation (5.8) leads to a nonlinear steepening of the $\bar{h}_0$-field in the zonal (x) direction which has a tendency to decrease the characteristic space scale. This phenomenon may violate the self-consistency of the FD model if the characteristic scale goes down to the Rossby scale R_d in finite time. To be precise, this catastrophic scenario is realized in the case of a one-dimensional profile of the initial $\bar{h}_0$:

$$\bar{h}_{0I} = \bar{h}_{0I}(ax + by). \tag{5.9}$$

In this case the profile remains one-dimensional forever, the Jacobian in (5.8) vanishes identically and it becomes a simple wave equation describing breaking in finite time, as is well-known. Let us stress, however, that initial condition (5.9) is not of the spatially localized (in both directions) form required by the present study. The steepening cannot be catastrophic for spatially localized $\bar{h}_{0I}$ of the form $\bar{h}_{0I} = 1 + h_{d0}$, where $h_{d0}(x, y)$ is a function decaying at infinity. It is known (cf. Cushman-Roisin 1986; Ben Jelloul & Zeitlin 1999) that there are infinitely many conserved quantities in the dynamics described by (5.8). They are the so-called Casimir invariants

$$C_F = \int \mathrm{d}x\,\mathrm{d}y F(h_d) = \mathrm{const} \tag{5.10}$$

and the following functional related to the kinetic energy (cf. (4.30)):

$$E = \int \mathrm{d}x\,\mathrm{d}y\left[(1 + h_d)(\nabla h_d)^2 - \bar{\beta}yh_d^2\right] = \mathrm{const}. \tag{5.11}$$

Here $F(h_d)$ is an arbitrary function of h_d and we replace $\bar{h}_0$ everywhere by $1+h_d(x, y; t)$. Let us show that the constraints imposed by these conservation laws prohibit breaking. Suppose, for simplicity, that h_d is a continuously differentiable function with compact support $\mathscr{D}$. It readily follows from (5.8) that h_d remains zero outside $\mathscr{D}$ for all times. Therefore, by virtue of (5.10) with $F(h_d) = h_d^2$ the value of $|\int \mathrm{d}x\,\mathrm{d}y\, yh_d^2|$ is bounded from above by some constant. Therefore, the positive-definite integral $\int \mathrm{d}x\,\mathrm{d}y(1 + h_d)(\nabla h_d)^2$ is bounded as well and the developement of infinite gradients of h_d (breaking) is forbidden. Hence, one may assume that for spatially localized disturbances the Jacobian terms in (5.8) can stop (or, at least, decelerate) the nonlinear steepening due to the β-term. Of course, this semi-qualitative consideration should be verified by a direct high-resolution numerical simulation, which is in progress and will be presented elsewhere.

The β-term in the amplitude equation (5.7) does not violate the energy conservation of the inertial oscillations (4.29). As is however clear from the standard quantum-mechanical-type analysis of this equation (e.g. for $\bar{h}_0 = \text{const}$), the β-term will produce a systematic meridional shift of the wave packet and a tendency to decrease the spatial scale of $\mathscr{A}$ in time. Thus, the model may lose self-consistency (the details depend on the initial spatial distribution of the inertial oscillations) on times $\sim \epsilon^{-2}$, in analogy with and for the same reasons as in the QG case. Thus the multiple-space-scales technique should be applied in order to avoid difficulties (cf. Pedlosky 1984).

6. Summary and discussion

To summarize, by applying multiple-time-scale perturbation theory we were able to describe the process of the geostrophic adjustment of localized disturbances in various regimes of the RSW dynamics at small Rossby numbers.

On the f-plane we find that an arbitrary perturbation is split into a unique way into slow and fast components evolving with characteristic time scales f_0^{-1} and $(f_0 Ro)^{-1}$ respectively. The slow component is not influenced by the fast one and remains close to the geostrophic balance. We show that the scenario of adjustment depends on the initial relative elevation of the free surface $\Delta H/H_0$, where ΔH and H_0 are the typical values of the initial elevation and the mean depth, respectively.

For $\Delta H/H_0 = O(Ro)$ the evolution of the slow motion is governed by the well-known quasi-geostrophic potential vorticity equation for times $t \leqslant (f_0 Ro)^{-1}$ and by the modified geostrophic potential vorticity equation for times $t \leqslant (f_0 Ro^2)^{-1}$. The fast component consists mainly of linear inertia–gravity waves rapidly propagating outward from the initial disturbance; their interactions with each other and with the slow component result in a small correction to the fast field only.

For $\Delta H/H_0 \gg Ro$ the slow field is governed by the frontal geostrophic dynamics equation when $\Delta H/H_0 = O(1)$ (frontal regime) or by its simplified version in the intermediate case $Ro \ll \Delta H/H_0 \ll 1$ (nonlinear geostrophic regime). The fast component in these cases is a spatially localized packet of almost non-dispersive inertial oscillations coupled to the slow component of the flow. Its envelope experiences a slow modulation and obeys a Schrödinger-type modulation equation describing advection and dispersion of the packet. Depending on the slow component profile the trapping of inertial oscillations may occur.

The physical reason for the slow–fast splitting, as could be anticipated and is seen in perturbative calculations, is the Lagrangian conservation of potential vorticity, the fact that inertia–gravity waves do not carry potential vorticity, and the gap in the spectrum of small perturbations in RSW due to rotation which, in particular, blocks the Lighthill radiation of inertia–gravity waves.

We thus establish the fast–slow motion splitting or, in other words, we demonstrate the existence of the 'perturbative' slow manifold in the RSW dynamics, the splitting being equivalent to 'non-acceleration', i.e. to the absence of wave drag of the slow motion. We therefore corroborate, by using an alternative technique and analysing the localized disturbances evolving in the infinite domain, the results obtained earlier in the discrete spectrum case by Embid & Majda (1996) and Babin *et al.* (1998*a*, *b*). There exists, however, an important difference between the continuous and discrete spectrum cases, the former being insensitive at the low orders of the perturbation theory to the fast–fast and slow–fast resonances diluted in the continuous spectrum, while these resonances are essential in the latter case.

We derive the evolution equations for the slow component of motion and present a well-defined initialization procedure which, unlike the standard initializations (Baer

& Tribbia 1977; Machenhauer 1977) does not presume the smallness of the fast component of motion, although has another restriction of smallness of the Rossby number. The resulting fast component (inertia–gravity waves and inertial oscillations) is completely resolved for a given set of initial conditions.

Of course, the perturbative character of our demonstration imposes obvious self-consistency restrictions and has a domain of validity limited in time. Thus, the motion should preserve its single-scale character for both components and there should not be explosive finite-time instabilities in slow dynamics (it is intuitively clear that these two requirements are not independent) destroying the slow-time scaling. The limits of the single-space-scale approach become evident while changing to the β-plane equations. Nevertheless, the splitting persists on the β-plane too, as we show.

Let us emphasize an important practical conclusion following from the fast–slow splitting established for the RSW equations: the evolution properties of slow structures, in general, and their stability, in particular, may be safely studied within the slow dynamics equations as they are not influenced at all by the fast component of the flow.

It is of importance that numerical simulations (Dewar & Killworth 1995; Jones, Mahalov & Nicolaenko 1999) of RSW dynamics confirm the fast–slow splitting and non-interaction of quasi-geostrophic and wave components of the flow, at least in doubly periodic domains (in order to have a reliable comparison with our results the sponge boundary conditions absorbing waves should be applied). Thus Dewar & Killworth (1995) have demonstrated non-interaction by analysing energy and enstrophy behaviour in their pseudo-spectral calculation but failed to reach quantitative agreement with corresponding QG dynamics, presumably because of improper initialization. Note that our demonstration of (3.32) is very close to their leading-order proof of non-interaction. Jones *et al.* (1999) also corroborated splitting by spectral calculations with initialization corresponding to that we used in (3.13), (3.14), (3.15), (3.17), (3.19). (Correspondingly, we believe that these results may be further improved by adjusting initialization and taking into account corrections to the QGPV equation as it was done in §§ 3.2 and 3.3 above.)

After the present paper had been submitted for publication, instructive numerical results of fully nonlinear Rossby adjustment (no smallness of Rossby number in initial data) were obtained by Kuo & Polvani (2000) indicating an influence of the inertia–gravity waves on the vortical component of the flow in this case.

G. M. R. gratefully acknowledges the kind hospitality of the LMD during his stay in Paris. His work was supported in part by RFBR grant No 99-05-64841. The work of V. Z. was supported in part by the EC grant No ENV4-CT97-0546 ('TRACAS'). We thank the participants of the Working group 'Gravity waves and stratified turbulence' of the IPSL (Paris) for their helpful comments We are also grateful to W. Blumen for numerous stimulating discussions and for indicating us some references of which we were unaware.

Appendix A. The stationary phase calculations

The stationary phase method is a standard tool in the theory of linear waves (cf. Witham 1974, chap. 11). It is applied in the calculations of large-time asymptotics of wave packets, although most frequently at $\boldsymbol{x}/t$ fixed. Our goal is the large-t asymptotics at *fixed* position. For an arbitrary wave packet of inertia–gravity waves of the form

$$\mathscr{F}(\boldsymbol{x};t) = \int \mathrm{d}\boldsymbol{k}\,\hat{\mathscr{F}}(\boldsymbol{k})\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}\pm\Omega_{\boldsymbol{k}}t)} \tag{A 1}$$

the stationary phase method in its standard form (cf. e.g. Olver 1974, chap. 3) may be applied as $\Omega_{\boldsymbol{k}}$ is a convex function with a single minimum at $|\boldsymbol{k}| = 0$. The main contribution at $t \to \infty$ is given by the stationary point of the phase with respect to the integration variables and, thus,

$$\mathscr{F}(\boldsymbol{x};t)|_{t\to\infty} = O(1/t) \tag{A 2}$$

(each integration giving a $t^{-1/2}$ contribution and the oscillating factor $\mathrm{e}^{\pm\mathrm{i}f_0 t}$, as usual). Alternatively, one can use the symmetry of $\Omega_{\boldsymbol{k}}$, i.e. its independence of the angular variable in $\boldsymbol{k}$-space. Then the above asymptotics is obtained by a straightforward integration by parts in the $|\boldsymbol{k}|$-integral (this method will be applied below).

Now consider solutions $h(\boldsymbol{x};t)$ of the inhomogeneous KG equation with the following right-hand side corresponding to the expressions arising at the third order of the perturbation theory in the QG regime (cf. § 3):

$$-\frac{\partial^2 h}{\partial t^2} - h + \nabla^2 h = \mathscr{F}^{(s)}(\boldsymbol{x})\,\mathscr{F}_0^{(f)}(\boldsymbol{x};t) + \mathscr{F}_1^{(f)}(\boldsymbol{x};t)\,\mathscr{F}_2^{(f)}(\boldsymbol{x};t) \equiv Q^{(sf)} + Q^{(ff)}, \tag{A 3}$$

where $\mathscr{F}^{(s)}, \mathscr{F}_{1,2}^{(f)}$ are spatially localized functions with well-defined Fourier-transforms:

$$\mathscr{F}^{(s)}(\boldsymbol{x}) = \int \mathrm{d}\boldsymbol{k}\,\hat{\mathscr{F}}^{s}(\boldsymbol{k})\,\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \quad \mathscr{F}_i^{(f)}(\boldsymbol{x};t) = \int \mathrm{d}\boldsymbol{k}\,\hat{\mathscr{F}}_i^{f}(\boldsymbol{k})\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}\pm\Omega_{\boldsymbol{k}}t)}, \quad i = 0, 1, 2 \tag{A 4}$$

and, hence,

$$Q^{(sf)} = \int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2\,\hat{\mathscr{F}}^{s}(\boldsymbol{k}_1)\,\hat{\mathscr{F}}_0^{f}(\boldsymbol{k}_2)\mathrm{e}^{\mathrm{i}((\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{x}\pm\Omega_{\boldsymbol{k}_2}t)}, \tag{A 5}$$

$$Q^{(ff)} = \int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2\,\hat{\mathscr{F}}_1^{f}(\boldsymbol{k}_1)\,\hat{\mathscr{F}}_2^{f}(\boldsymbol{k}_2)\mathrm{e}^{\mathrm{i}((\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{x}+(\pm\Omega_{\boldsymbol{k}_1}\pm\Omega_{\boldsymbol{k}_2})t)}. \tag{A 6}$$

Using the superposition principle we consider separately the fast–fast and the slow–fast contributions on the right-hand side. Let us start with the simpler case of $Q^{(ff)}$. The corresponding solution of (A 3) is immediately written as

$$h^{(ff)}(\boldsymbol{x};t) = \int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2\,\mathrm{e}^{\mathrm{i}((\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{x}+(\pm\Omega_{\boldsymbol{k}_1}\pm\Omega_{\boldsymbol{k}_2})t)}\,\frac{\hat{\mathscr{F}}_1^{f}(\boldsymbol{k}_1)\hat{\mathscr{F}}_2^{f}(\boldsymbol{k}_2)}{\Omega^2_{\boldsymbol{k}_1+\boldsymbol{k}_2} - (\Omega^2_{\boldsymbol{k}_1} \pm \Omega^2_{\boldsymbol{k}_2})^2}. \tag{A 7}$$

There exist no resonant triads of inertia–gravity waves, therefore the denominator of this expression never vanishes. Hence, the standard stationary phase method may be safely applied giving

$$h^{(ff)}(\boldsymbol{x};t)|_{t\to\infty} = O(1/t^2) \tag{A 8}$$

at a fixed point $\boldsymbol{x}$.

The analysis of the slow–fast solution $h^{(sf)}$ in (A 3) is more subtle because the integrand of $h^{(sf)}$ does contain singularities

$$h^{(sf)}(\boldsymbol{x};t) = \int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2\,\mathrm{e}^{\mathrm{i}((\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{x}\pm\Omega_{\boldsymbol{k}_2}t)}\,\frac{\hat{\mathscr{F}}^{s}(\boldsymbol{k}_1)\hat{\mathscr{F}}_0^{f}(\boldsymbol{k}_2)}{\Omega^2_{\boldsymbol{k}_1+\boldsymbol{k}_2} - \Omega^2_{\boldsymbol{k}_1}} \tag{A 9}$$

corresponding to wave–wave–mean flow resonance.

To estimate $h^{(sf)}$ we introduce new integration variables $\boldsymbol{k} = \boldsymbol{k}_1 + \boldsymbol{k}_2, \boldsymbol{k}_2$ and rewrite

$Q^{(sf)}$ using the polar coordinates in Fourier space:

$$Q^{(sf)}(\boldsymbol{x},t) = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} \mathrm{d}k\, k \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \int_0^{2\pi} \mathrm{d}\phi_2 \int_0^{\infty} \mathrm{d}k_2 \hat{\mathscr{F}} \mathrm{e}^{\pm \mathrm{i}\Omega_{k_2} t}. \tag{A 10}$$

Here

$$\hat{\mathscr{F}} = \hat{\mathscr{F}}(\boldsymbol{k},\boldsymbol{k}_2) = k_2 \hat{\mathscr{F}}^s(\boldsymbol{k}-\boldsymbol{k}_2)\hat{\mathscr{F}}^f(\boldsymbol{k}_2) \tag{A 11}$$

and $\Omega_{\boldsymbol{k}} \equiv \Omega(k) = \sqrt{k^2+1}$. Using a new variable $u = k_2 - k$ we have

$$Q^{(sf)}(\boldsymbol{x},t) = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} \mathrm{d}k\, k \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \int_0^{2\pi} \mathrm{d}\phi_2 \int_{-k}^{\infty} \mathrm{d}u \hat{\mathscr{F}} \mathrm{e}^{\pm \mathrm{i}\Omega(u+k)t}. \tag{A 12}$$

To isolate the 'resonant' point $u = 0$ we divide the integration domain $(-k,\infty)$ into three sub-domains: $\mathscr{I}_0 = [-\min(k,\epsilon),\epsilon], \mathscr{I}_1 = [\epsilon,\infty], \mathscr{I}_2 = [-k,-\min(k,\epsilon)]$ where $\epsilon \to 0$, and represent $Q^{(sf)}$ in the form $Q^{(sf)} = Q_0 + Q_1 + Q_2$, correspondingly. The function $h^{(sf)}$ is written as a sum

$$h^{(sf)} = h_0^{(sf)} + h_1^{(sf)} + h_2^{(sf)}, \tag{A 13}$$

where $h_i^{(sf)}$, $i = 0,1,2$, satisfy the equations

$$-\frac{\partial^2 h_i^{sf}}{\partial t^2} - h_i^{sf} + \nabla^2 h_i^{sf} = Q_i, \quad i = 0,1,2. \tag{A 14}$$

The source terms Q_1, Q_2 do not contain resonances and, therefore,

$$h_i^{(sf)}(\boldsymbol{x};t)|_{t\to\infty} = O(1/t^2), \quad i = 1,2, \tag{A 15}$$

at a fixed point $\boldsymbol{x}$ like the function $h^{(ff)}(\boldsymbol{x},t)$ above.

Using the smallness of ϵ the solution $h_0^{(sf)}$ can be represented approximately in the form

$$h_0^{(sf)}(\boldsymbol{x},t) = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} \mathrm{d}k\, k \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} \pm \Omega_{\boldsymbol{k}} t)} \hat{\mathscr{G}}(k,\phi) \int_{-\min(k,\epsilon)}^{\epsilon} \mathrm{d}u\, s(u,k,t), \tag{A 16}$$

where

$$\hat{\mathscr{G}} = \frac{1}{2\Omega_{\boldsymbol{k}}} \int_0^{2\pi} \mathrm{d}\phi_2 \hat{\mathscr{F}}(\boldsymbol{k},\boldsymbol{k}_2)|_{k_2=k}, \quad s(u,k,t) = \frac{1-\mathrm{e}^{\pm \mathrm{i}[\Omega(k+u)-\Omega(k)]t}}{\Omega(k)-\Omega(k+u)}. \tag{A 17}$$

The function $h_0^{(sf)}$ in (A 16) can be represented as a sum

$$h_0^{(sf)} = h_{01}^{(sf)} + h_{02}^{(sf)} \tag{A 18}$$

where

$$h_{01}^{(sf)}(\boldsymbol{x},t) = \int_0^{2\pi} \mathrm{d}\phi \int_{\epsilon}^{\infty} \mathrm{d}k\, k \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} \pm \Omega_{\boldsymbol{k}} t)} \hat{\mathscr{G}}(k,\phi) \int_{-\epsilon}^{\epsilon} \mathrm{d}u\, s(u,k,t) \tag{A 19}$$

and

$$h_{02}^{(sf)}(\boldsymbol{x},t) = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\epsilon} \mathrm{d}k\, k \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} \pm \Omega_{\boldsymbol{k}} t)} \hat{\mathscr{G}}(k,\phi) \int_{-k}^{\epsilon} \mathrm{d}u\, s(u,k,t). \tag{A 20}$$

Since

$$\int_{-\epsilon}^{\epsilon} \mathrm{d}u\, s(u,k,t) = \pm \frac{\mathrm{i}\,\pi}{\Omega'(k)} + O\left(\frac{1}{t}\right), \quad t \to \infty, \quad k > 0 \tag{A 21}$$

we get

$$h_{01}^{(sf)}(\boldsymbol{x};t)|_{t\to\infty} = O(1/t). \tag{A 22}$$

When estimating $h_{02}^{(sf)}$ we use the fact that u and k are small in (A 20) due to the smallness of ϵ and, therefore, the functions $\Omega(k)$ and $s(u,k,t)$ in the integrals on the right-hand side of (A 20) can be simplified. Rather tedious calculations show that the function $h_{02}^{(sf)}$ has the same asymptotics at $t \to \infty$ as $h_{01}^{(sf)}$ and, hence,

$$h^{(sf)}(\boldsymbol{x};t)|_{t\to\infty} = O(1/t) \tag{A 23}$$

at a fixed point $\boldsymbol{x}$.

Therefore the full solution of (A 3) $h^{(sf)}+h^{(ff)}$ is well-defined and displays no secular growth. In other words, although a slow–fast resonance is formally present in (A 9) its measure is zero in the continuous spectrum and gives no dangerous contributions.

Let us make a brief remark concerning the higher orders of the perturbation theory. Higher-order wave–wave resonances (four-wave etc.) will appear due to the fast–fast source term. Although they will engender the singularities of the denominators in the integrands of the solutions expressed with the help of the Fourier integrals, these singularities are, presumably, integrable as was the case above for the wave–mean resonance and, thus, do not require introduction of slow time-dependence in the wave amplitudes. Another possible source of singularities comes from the slow–fast source term where the function $\mathscr{F}^s$ will contain contributions from the KG solutions integrated in time (cf. (3.35) and below). The time integration (cf. (A 1)) makes $\Omega_{\boldsymbol{k}}$ appear in the denominator in the Fourier-integrals. However, due to the gap in the spectrum (2.10) this procedure is safe and, even multiply repeated, cannot lead to small-$\boldsymbol{k}$ ('infrared') singularities. Physically speaking, this reflects a Lighthill radiation (Lighthill 1952) supression in the presence of the spectral gap. This is not the case in the absence of rotation which, thus, plays a crucial rôle. Finally, let us remark that linear resonances will also appear at the higher orders. They are, however, insignificant and may be treated like those above in (3.53).

Appendix B. The details of the NLQG calculations

B.1. $O(\lambda^0)$ solution

In complex notation

$$\partial_t \mathscr{U}_0 + \mathrm{i}\,\mathscr{U}_0 = -2\partial_{z^*} h_0 \tag{B 1}$$

and

$$\partial_t h_0 = 0. \tag{B 2}$$

The solution is then

$$\left.\begin{aligned} h_0 &= h_0(x,y;t_1,t_2,\ldots),\\ \mathscr{U}_0 &= \mathscr{A}_0(t_1,t_2,\ldots)\mathrm{e}^{-\mathrm{i}t} + 2\mathrm{i}\partial_{z^*} h_0(x,y;t_1,\ldots). \end{aligned}\right\} \tag{B 3}$$

The height field does not depend on the fastest time scale and the velocity field is a superposition of a slowly evolving geostrophic flow and a purely inertial oscillations.

B.2. $O(\lambda^1)$ solution

$$\partial_t \mathscr{U}_1 + \mathrm{i}\,\mathscr{U}_1 = -(\partial_{t_1}\mathscr{U}_0 + 2\partial_{z^*} h_1), \tag{B 4}$$

$$\partial_t h_1 = -\partial_{t_1} h_0 - \partial_z \mathscr{U}_0 - \partial_{z^*}\mathscr{U}_0^*. \tag{B 5}$$

Integration over the fast time t of the height equation gives $\partial_{t_1} h_0 = 0$ and the fast–slow decomposition results:

$$h_1 = \bar{h}_1(t_1,\ldots) + [\mathscr{H}_1^-(t_1,\ldots)\mathrm{e}^{-\mathrm{i}t} + \mathrm{c.c.}], \tag{B 6}$$

where $\mathscr{H}_1^-$ may be easily found:

$$\mathscr{H}_1^- = -\mathrm{i}\partial_z \mathscr{A}_0. \tag{B 7}$$

Eliminating the resonant terms on the right-hand side of (B 4) gives the amplitude equation

$$\partial_{t_1} \mathscr{A}_0 - 2\mathrm{i}\partial_{zz^*} \mathscr{A}_0 = 0. \tag{B 8}$$

The full expression for the first-order perturbation is

$$\mathscr{U}_1 = \mathscr{A}_1^- \mathrm{e}^{-\mathrm{i}t} + \bar{\mathscr{U}}_1 + \mathscr{A}_1^+ \mathrm{e}^{\mathrm{i}t} \tag{B 9}$$

where

$$\mathscr{A}_1^- = \mathscr{A}_1^-(t_1, \ldots), \qquad \bar{\mathscr{U}}_1 = 2\mathrm{i}\partial_{z^*} \bar{h}_1, \qquad \mathscr{A}_1^+ = \mathrm{i}\partial_{z^*} \mathscr{H}_1^{-^*}, \tag{B 10}$$

B.3. *$O(\lambda^1)$ solution*

$$\left.\begin{aligned}\partial_t \mathscr{U}_2 + \mathrm{i}\,\mathscr{U}_2 &= -[\partial_{t_2}\mathscr{U}_0 + \partial_{t_1}\mathscr{U}_1 + (\mathscr{U}_0\partial_z + \mathscr{U}_0^*\partial_{z^*})\mathscr{U}_0 + 2\partial_{z^*}h_2],\\ \partial_t h_2 &= -[\partial_{t_2}h_0 + \partial_{t_1}h_1 + \partial_z(h_0\mathscr{U}_0) + \partial_{z^*}(h_0\mathscr{U}_0^*) + \partial_z\mathscr{U}_1 + \partial_{z^*}\mathscr{U}_1^*].\end{aligned}\right\} \tag{B 11}$$

Averaging (B 11) over the fast time yields

$$\partial_{t_2} h_0 + \partial_{t_1} \bar{h}_1 = \partial_z \bar{\mathscr{U}}_1 + \partial_{z^*} \bar{\mathscr{U}}_1^* + \langle \partial_z (h_0\,\mathscr{U}_0) + \partial_{z^*}(h_0\,\mathscr{U}_0^*) \rangle = 0. \tag{B 12}$$

To avoid resonances we put

$$\partial_{t_2} h_0 = 0, \qquad \partial_{t_1} \bar{h}_1 = 0. \tag{B 13}$$

Eliminating the resonance in the velocity equation (B 11) yields

$$\partial_{t_1} \mathscr{A}_1^- + 2\partial_{z^*} \mathscr{H}_2^- + \mathscr{A}_0 \partial_z \bar{\mathscr{U}}_0 + \partial_{t_2} \mathscr{A}_0 + \bar{\mathscr{U}}_0 \partial_z \mathscr{A}_0 + \bar{\mathscr{U}}_0^* \partial_{z^*} \mathscr{A}_0 = 0 \tag{B 14}$$

with

$$\mathscr{H}_2^- = -\mathrm{i}[\partial_{t_1} \mathscr{H}_1^- + \partial_z (h_0 \mathscr{A}_0) + \partial_z \mathscr{A}_1^- + \partial_{z^*} \mathscr{A}_1^{+*}]. \tag{B 15}$$

We finally get the evolution equation for $\mathscr{A}_1$

$$\begin{aligned}\partial_{t_1} \mathscr{A}_1 - 2\mathrm{i}\partial^2_{zz^*} \mathscr{A}_1 = -[&\partial_{t_2} \mathscr{A}_0 + 2\mathrm{i}\mathscr{A}_0 \partial^2_{zz^*} h_0 + J(h_0, \mathscr{A}_0)\\ &-2\partial_{t_1}\partial^2_{zz^*} \mathscr{A}_0 + 2\mathrm{i}\partial^4_{z^*z^*zz} \mathscr{A}_0 - 2\mathrm{i}\partial^2_{zz^*} h_0 \mathscr{A}_0].\end{aligned} \tag{B 16}$$

Using the full variables

$$h = 1 + \lambda h_0, \qquad \mathscr{A} = \mathscr{A}_0 + \lambda \mathscr{A}_1, \qquad \partial_\tau = \partial_{t_1} + \lambda \partial_{t_2}. \tag{B 17}$$

one obtains an 'improved' amplitude equation

$$\partial_t \left(\mathscr{A} - \frac{\lambda}{2} \nabla^2 \mathscr{A} \right) - \frac{\mathrm{i}}{2} \nabla^2 (h \mathscr{A}) + \lambda \left[J(h, \mathscr{A}) + \mathrm{i} \mathscr{A} \left(\frac{\nabla^2 h}{2} \right) + \frac{\mathrm{i} \nabla^4 \mathscr{A}}{8} \right] = O(\lambda^2). \tag{B 18}$$

The full expression of the perturbation velocity field $\mathscr{U}_2$ is

$$\mathscr{U}_2 = \mathscr{A}_2^{--} \mathrm{e}^{-2\mathrm{i}t} + \mathscr{A}_2^- \mathrm{e}^{-\mathrm{i}t} + \bar{\mathscr{U}}_2 + \mathscr{A}_1^+ \mathrm{e}^{\mathrm{i}t} \tag{B 19}$$

where the slowly varying amplitudes may be easily found:

$$\left.\begin{aligned}\mathscr{A}_2^{--} &= -\mathrm{i}\mathscr{A}_0 \partial_z \mathscr{A}_0,\\ \bar{\mathscr{U}}_2 &= \mathrm{i}\mathscr{A}_0^* \partial_{z^*} \mathscr{A}_0 - 2J(h_0, \partial_{z^*} h_0) + 2\mathrm{i}\partial_{z^*} \bar{h}_2,\\ \mathscr{A}_2^+ &= -\mathscr{A}_0^* \partial^2_{z^*z^*} h_0 + \mathrm{i}\partial_{z^*} \mathscr{H}_2^{-^*} - \tfrac{1}{4} \partial^2_{zz^*} \mathscr{A}_0^+.\end{aligned}\right\} \tag{B 20}$$

B.4. *Order-three solutions*

$$\partial_t \mathscr{U}_3 + \mathrm{i}\,\mathscr{U}_3 = -[\partial_{t_1}\mathscr{U}_2 + \partial_{t_2}\mathscr{U}_1 + \partial_{t_3}\mathscr{U}_0 + (\mathscr{U}_0\partial_z + \mathscr{U}_0^*\partial_{z^*})\mathscr{U}_1 + (\mathscr{U}_1\partial_z + \mathscr{U}_1^*\partial_{z^*})\mathscr{U}_0 + 2\partial_{z^*}h_3], \quad (\mathrm{B}\,21)$$

$$\partial_t h_3 = -[\partial_{t_3}h_0 + \partial_{t_1}h_2 + \partial_{t_2}h_1 + \partial_z(h_0\mathscr{U}_1) + \partial_{z^*}(h_0\mathscr{U}_1^*) + \partial_z h_1\mathscr{U}_0 + \partial_{z^*}h_1\mathscr{U}_0^* + \partial_z\mathscr{U}_2 + \partial_{z^*}\mathscr{U}_2^*]. \quad (\mathrm{B}\,22)$$

The height field equation is obtained by averaging over times t, t_1 and t_2. One finds that the quadratic contributions from the wave field exactly vanish

$$\mathrm{i}\partial_z\mathscr{A}_0\partial_{z^*}\mathscr{A}_0^* + \mathrm{i}\partial_z\mathscr{A}_0^*\partial_{z^*}\mathscr{A}_0 + \text{c.c.} = 0. \quad (\mathrm{B}\,23)$$

Eliminating resonances gives the NLQG equation (2.16) for h_0.

REFERENCES

Allen, J. S. 1993 Iterated geostrophic intermediate models. *J. Phys. Oceanogr.* **23**, 2447–2461.

Babin, A., Mahalov, A. & Nicolaenko, B. 1998*a* Global splitting and regularity of rotating shallow-water equations. *Eur. J. Mech.* B/*Fluids* **15**, 291–300.

Babin, A., Mahalov, A. & Nicolaenko, B. 1998*b* Regularity and integrability of rotating shallow-water equations. *C. R. Acad. Sci. Paris* I **324**, 593–598.

Baer, F. & Tribbia, J. 1977 On complete filtering of gravity modes through nonlinear initialization. *Mon. Wea. Rev.* **105**, 1536–1539.

Ben Jelloul, M. & Zeitlin, V. 1999 Remarks on stability of the rotating shallow water vortices in the frontal regime. *Nuovo Cim.* C **22**, 931–941.

Blumen, W. 1972 Geostrophic adjustment. *Rev. Geophys. Space Phys.* **10**, 485–528.

Blumen, W. & Wu, R. 1995 Geostrophic adjustment: Frontogenesis and energy conversion. *J. Phys. Oceanogr.* **25**, 428–438.

Charney, J. G. 1948 On the scale of atmospheric motion. *Geofys. Publ. Oslo* **17**, 1–17.

Charney, J. G. & Flierl, G. R. 1981 Oceanic analogs of large-scale atmospheric motions. In *Evolution of Physical Oceanography* (ed. B. A. Warren & C. Wunsch), pp. 504–548. MIT Press

Cullen, M. J. P. 1998 The use of dynamical knowledge of the atmosphere to improve NWP models. *British Meterological Office Rep.* 52.

Cushman-Roisin, B. 1986 Frontal geostrophic dynamics. *J. Phys. Oceanogr.* **16**, 132–143.

Dewar, W. K. & Killworth, P. D. 1995 Do fast gravity waves interact with geostrophic motions? *Deep-Sea Res.* I **42**, 1063–1081.

Embid, P. F. & Majda, A. J. 1996 Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity. *Commun. Part. Diffl Equat.* **21**, 619–658.

Fal'kovich, G. E. 1992 Inverse cascade and wave condensate in mesoscale atmospheric turbulence. *Phys. Rev. Lett.* **69**, 3173–3176.

Fal'kovich, G. E., Kuznetsov, E. & Medvedev, S. B. 1994 Nonlinear interaction between long inertio-gravity and Rossby waves. *Nonlin. Proc. Geophys.* **1**, 168–172.

Fal'kovich, G. E. & Medvedev, S. B. 1992 Kolmogorov-like spectrum for turbulence of inertial-gravity waves. *Europhys. Lett.* **19**, 279–284.

Gill, A. E. 1982 *Atmosphere-Ocean Dynamics*, Chap. 7.2. Academic.

Holton, J. R. 1979 *An Introduction to Dynamic Meteorology*. Academic.

Jones, D. A., Mahalov, A. & Nicolaenko, B. 1999 A numerical study of an operator splitting method for rotating flows with large ageostrophic initial data. *Theoret. Comput. Fluid Dyn.* **13**, 143–159.

Kuo, A. C. & Polvani, L. M. 2000 Nonlinear geostrophic adjustment, cyclone/anticyclone asymmetry, and potential vorticity rearrangement. *Phys. Fluids* **12**, 1087–1100.

Lighthill, M. J. 1952 On sound generated aerodynamically. I. General theory. *Proc. R. Soc. Lond.* A **211**, 564–587.

Machenhauer, B. 1977 On the dynamics of gravity oscillations in a shallow water model, with application to normal-mode initialization. *Contrib. Atmos. Phys.* **50**, 253–271.

Medvedev, S. B. 1999 The slow manifold for the shallow water equations on the f-plane. *J. Atmos. Sci.* **56**, 1050–1054.

Monin, A. S. & Obukhov, A. M. 1958 Small oscillations of the atmosphere and adjustment of meteorological fields. *Izv. Geophys.* **11**, 1360–1373 (in Russian).

Obukhov, A. M. 1949 On the problem of geostrophic wind. *Izv. Geogr. Geophys.* **13**, 281–306 (in Russian).

Olver, F. W. J. 1974 *Introduction to Asymptotics and Special Functions*, Chap. 3. Academic.

Pedlosky, J. 1982 *Geophysical Fluid Dynamics*. Springer.

Pedlosky, J. 1984 The equations for geostrophic motion in the ocean. *J. Phys. Oceanogr.* **14**, 448–455.

Reznik, G. M. & Grimshaw, R. 2001 Ageostrophic dynamics of an intense localized vortex on the β-plane. *J. Fluid Mech.* **443**, 351–376.

Reznik, G. M., Grimshaw, R. & Benilov, E. S. 2000 On the long-term evolution of an intense localised divergent vortex on the β-plane. *J. Fluid Mech.* **422**, 249–280.

Romanova, N. N. & Tseitlin (Zeitlin), V. 1984 On quasigeostrophic motions in barotropic and baroclinic fluids. *Izv. Atmos. Ocean Phys.* **20**, 85–91.

Rossby, C.-G. 1938 On the mutual adjustment of pressure and velocity distributions in certain simple current systems. *J. Mar. Res.* **1**, 239–263.

Stegner, A. & Zeitlin, V. 1995 What can asymptotic expansions tell us about large-scale quasi-geostophic anticyclonic vortices? *Nonlin. Proc. Geophys.* **2**, 186–193.

Stegner, A. & Zeitlin, V. 1998 From quasi-geostrophic to strongly nonlinear monopolar vortices in paraboloidal shallow-water-layer experiment. *J. Fluid Mech.* **356**, 1–24.

Warn, T., Bokhove, O., Shepherd, T. G. & Vallis, G. K. 1995 Rossby number expansions, slaving principles, and balanced dynamics. *Q. J. R. Met. Soc.* **121**, 723–739.

Williams, G. P. & Yamagata, T. 1984 Geostrophic regimes, intermediate solitary vortices and Jovian eddies. *J. Atmos. Sci.* **41**, 453–478.

Whitham, G. B. 1974 *Linear and Nonlinear Waves*, Chap. 11. Wiley.

Young, W. R. & Ben Jelloul M. 1997 Propagation of near-inertial oscillations through a geostrophic flow. *J. Mar. Res.* **55**, 735–766.