6M.4 NON-MODAL GROWTHS OF SYMMETRIC PERTURBATIONS PRODUCED BY PAIRED NORMAL MODES

Qin Xu*
NOAA/National Severe Storms Laboratory, Norman, Oklahoma

1. Introduction

Symmetric instability (SI) was studied theoretically as one of the possible mechanisms for the formations of frontal rainbands (Bennetts and Hoskins 1979; Emanuel 1979, 1983; Xu and Zhou 1982; Miller 1985; Xu 1986, 2004). In the past, the SI theory was developed mainly in terms of modal growths and structures. The fastest growing mode, however, does not necessarily account for all the structures observed in unstable flows (Trefethen et al. 1993). When the modes are not orthogonal, certain linear combinations of two or more modes can grow faster than any individual mode over physically relevant time scales, and the combined structures are not time invariant or, say, not modal (Farrell 1984; Buizza and Palmer 1995; Farrell and Ioannou 1996). Non-modal structures and growths have been studied for baroclinic waves, but not much research has been done on non-modal growths of symmetric perturbations, so there is a gap between the classic SI and the non-modal growths of symmetric perturbations. An effort is made in this paper to fill this gap.

2. Governing equations and normal modes

a. Modal solutions

The basic state has an uniform stratification N^2 and an uniform vertical (thermal-wind) shear V_z (> 0). The inviscid instabilities of this basic state is controlled by two external parameters:

$$Ri = N^2/V_z \quad \text{Richardson number,} \tag{2.1a}$$
$$r = f/N \quad \text{inertial-buoyancy frequency ratio.} \tag{2.1b}$$

Denote by H the depth of the domain, then HV_z is used for the horizontal velocity scale, Hf for the vertical velocity scale, $1/f$ for the time scale, $L = HV_z/f$ for the horizontal length scale which is the Rossby radius of deformation associated with the basic shear.

Symmetric perturbations characterized by banded structures along the basic shear are governed by the following set of nondimensional equations and boundary conditions:

$$\Delta_a \psi_t + Ri\ b_x - v_z = 0, \tag{2.2a}$$
$$v_t + \psi_z - \psi_x = 0, \tag{2.2b}$$
$$b_t + \psi_z/Ri - \psi_x = 0. \tag{2.2c}$$
$$\psi = \psi_z = 0 \quad \text{at } z = 0, 1. \tag{2.3}$$

* *Corresponding author address:* Qin Xu, NSSL, 1313 Halley Circle, Norman, OK 73069;
e-mail: Qin.Xu@noaa.gov.

Here, ψ is the streamfunction, v the along-band perturbation velocity, and b the perturbation buoyancy, $\Delta_a = a^2(\ )_{xx} + (\ )_{zz}$, the subscripts $(\ )_{t,x,z}$ denote their respective partial differentials, and $a = H/L = r\sqrt{Ri}$ is the aspect ratio. In this paper, Ri ranges from 0.25 to 1.5, while r^2 is chosen to be 0.02. Since a is small, the model is close the hydrostatic limit ($a \to 0$) and the solution is insensitive to a.

The normal modes have the following form

$$\psi = \sin(n\pi z) \exp[\sigma t - ik(x + \beta z)], \tag{2.4}$$

where $\beta = (1 + \sigma^2)^{-1}$ and n is the (integer) vertical wavenumber. The normal mode expressions for v and b can be obtained by substituting (4) into (2b) and (2c), respectively. Substituting the resulting expressions with (4) into (2a) gives the following eigen-equation:

$$A\sigma^4 + 2B\sigma^2 + C = 0, \tag{2.5}$$

where $A = 1 + \mu^2 a^2$, $B = 1 + \mu^2(Ri + a^2)/2$, $C = 1 + \mu^2(Ri - 1)$, and $\mu = k/(n\pi)$. The roots of (2.5) are $\sigma_\pm^2 = (-B \pm \sqrt{D})/A$ where $D = B^2 - AC$. Note that $D = \mu^4(Ri - a^2)^2/4 + \mu^2(\mu^2 a^2 + 1) > (A - B)^2 = \mu^4(Ri - a^2)^2/4 \geq 0$. Since $D > 0$, both σ_+^2 and σ_-^2 are real, which means that the eigenvalue σ is either real or imaginary. Since $\sqrt{D} > |A - B|$, we have $\sigma_\pm^2 = -1 \pm \sqrt{D}/A + (A - B)/A$ and thus

$$\sigma_+^2 > -1 > \sigma_-^2. \tag{2.6}$$

Note also that $\sigma_+^2 > 0$ requires $D - B^2 = -AC > 0$ or, equivalently, $C < 0$ (since $A > 0$), which yields $1 \leq n^2 < (k/\pi)^2(1 - Ri)$. This means that there exist n_c growing modes (with $n = 1, 2, \dots n_c$) if

$$Ri < Ri_{nc} = 1 - (n_c l/2)^2, \tag{2.7}$$

where $l = 2\pi/k$ is the horizontal wavelength. Thus, $\sigma_+^2 > 0$ requires, at least, $n_c = 1$ in (2.7), that is, $Ri < Ri_c = 1 - (l/2)^2$.

The above solutions are insensitive to r as long as $r << 1$. When r^2 is small (or fixed), the squared roots σ_+^2 and σ_-^2 of (2.5) depend mainly (or solely) on Ri and μ. The squared root σ_+^2 is plotted as a function of (nl, Ri) for $r^2 = 0.02$ in Fig. 1, where $nl = 2n\pi/k = 2\mu^{-1}$. The mode structures will be examined for selected values of Ri and l. For this purpose, four parameter points are selected and marked in Fig. 1; that is, (l, Ri) = (0.2, 0.4) for case 1, (1.0, 0.7) for case 2, (1.5, 0.5) for case 3, and (0.1, 1.1) for case 4.

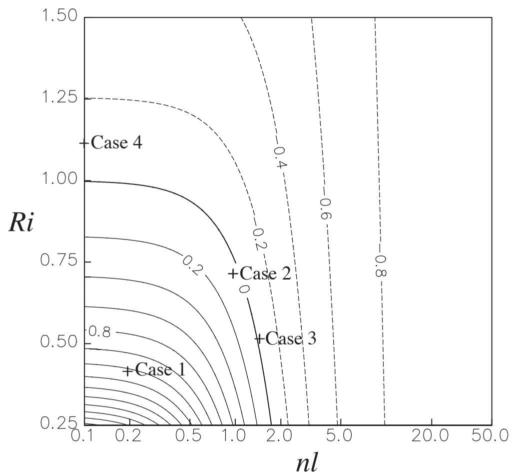

Fig. 1. $□_+^2$ plotted as a function of (nl, Ri) for r^2 = 0.02. Here l is the horizontal wavelength and $nl = 2□^{-1}$. Contours are every 0.2 with solid for non-negative and dashed for negative. The + signs mark the parameter points (with n = 1) for four cases.

b. Mode classification and polarization relationships

For given a, Ri and k, the two pairs of roots of (2.5) depend on n only and can be denoted by $\pm□_+(n)$ and $\pm□_-(n)$. Paired non-propagating growing and decaying modes exist for $\pm□_+(n)$ if $n □ n_c$ [see (2.7)]. In this case, the largest growth rate is given by $□_{max} = □_+(1)$ and the smallest positive growth rate is given by $□_{min} = □_+(n_c)$. When $n^2 > n_c$, $\pm□_+(n)$ become imaginary and the two paired modes become neutral and propagate in opposite directions. These modes are called slowly propagating modes. The slowest propagating modes are associated with $\pm□_+(n_s)$ where $n_s = n_c + 1$. Since $□_-^2$ is always negative, $\pm□_-$ are imaginary and the two paired modes are neutral and propagate with the same phase speed but in opposite directions. These modes [associated with $\pm□_-(n)$] are called fast propagating modes. The fastest propagating modes are associated with the gravest vertical mode of n = 1 and their phase speeds are given by $\pm i□_-(1)/k$.

By substituting the solutions in (2.4) back into (2.2b) and (2.2c), we obtain the following normal modes:

$$□ = \sin(n□z) \exp[□t - ik(x + □z)], \qquad (2.8a)$$
$$v = □^{-1}[ik(□ - 1)\sin(n□z) - n□ \cos(n□z)] \exp[□t - ik(x + □z)], \qquad (2.8b)$$
$$b = (□Ri)^{-1}[ik(□ - Ri)\sin(n□z) - n□ \cos(n□z)] \exp[□t - ik(x + □z)]. \qquad (2.8c)$$

Here, the amplitude of □ is set to unity in (2.8a) to facilitate the later comparisons between different modes. The normal modes in (2.8) are controlled by two external parameters (r, Ri) and two internal parameters (k, n). For a given set of values of these control parameters, there are four modes associated with the four roots of (2.5), that is, $\pm□_\pm$. According to (2.8a)-(2.8c), the two modes associated with each pair of $\pm□_-(n)$ or $\pm□_+(n)$ (□ 0) have exactly opposite polarization relationships between □ and (v, b). This means that the two modes have the same structure in □ but the opposite structures in (v, b) at the initial time (t = 0).

c. Stationary modes

If Ri □ Ri_{nc} in (2.7), then C □ 0 in (2.5). Consequently, □ □ $□_+(n_c) = 0$ and □ □ 1. In this case, the polarization relationship of (v, b) in (2.8b,c) with respect to □ in (2.8a) becomes singular. To keep (v, b) finite, we need to multiply □ to the solution in (2.8) and then let □ □ 0. In this case, the solution reduces to

$$□ = 0, \qquad (2.9a)$$
$$v = -n_c□ \cos(n_c□z) \exp[-ik(x + z)], \qquad (2.9b)$$
$$b = -n_c□Ri^{-1}[(i□_c)^{-1}\sin(n_c□z) + \cos(n_c□z)] \exp[-ik(x + z)], \qquad (2.9c)$$

where the condition of $□ - Ri = 1 - Ri = □_c^{-2} = (n_c□/k)^2$ (that is, C = 0) is used. This stationary mode satisfies (2.7) with n_c = 1 and is consistent with the solution in (2.8) in the limit of □ □ 0. It is easy to see that v and b in (2.9b,c) satisfy the thermal-wind relationship.

In addition to (2.9), there is another solution, which can be obtained from (2.8) - (2.9)/□ in the limit of □ □ 0 and thus has the following form:

$$□ = \sin(n_c□z) \exp[-ik(x + z)], \qquad (2.10a)$$
$$v = -t\, n_c□ \cos(n_c□z) \exp[-ik(x + z)], \qquad (2.10b)$$
$$b = -t\, n_c□Ri^{-1}[(i□_c)^{-1}\sin(n_c□z) + \cos(n_c□z)] \exp[-ik(x + z)], \qquad (2.10c)$$

where the Taylor expansion $\exp(□t) = 1 + □t + ...$ is used before taking the limit of □ □ 0. This solution is a linearly growing mode which satisfies (2.7). It is easy to see that v and b in (2.10b,c) satisfy the thermal-wind relationship while they grow linearly with time.

3. Mode structure analyses

a. Formulations for mode structure analyses

The nondimensional system of (2.2a-c) can be rewritten into the following form:

$$□_t = G, \qquad (3.1a)$$
$$v_t = J(□, M), \qquad (3.1b)$$
$$b_t = J(□, B), \qquad (3.1c)$$

where $□ = u_z - a^2 w_x = □_a□$ and $J(\ ,\ ) = (\ )_x(\ )_z - (\ )_z(\ )_x$ is the Jacobian bracket. Here, $M = x + z$ is the basic-state along-band absolute-momentum, that is, $fx + V$ scaled by fL, while $B = x/Ri + z$ + constant is the basic-state buoyancy, that is, $g□/□_o$ scaled by HN^2.

The vorticity generation term G in (3.1a) represents a torque imbalance (positive for a clockwise rotation) between the buoyancy torque $-Rib_x$ and the inertial-force torque v_z, where v represents the cross-band Coriolis force associated with the along-band velocity. In (3.1b), $J(□, M)$ represents the generation of v due to the cross-band advection of the basic-state M. In (3.1c), $J(□, B)$ and $-vB_y$ represent the generations of b due to the cross-band advection and along-band advection of the basic-state B, respectively. The slope of M-surface is -1 and the slope of B-surface is $-Ri^{-1}$ in the cross-band vertical plane (x, z). It is easy to see that $J(□, M)$ is zero when the streamline is parallel to the M-surface, and is positive (or negative) when the streamline is downward and steeper (or less steep) than the M-surface or when the streamline is upward and less steep (or steeper) than the M-surface in the cross-band vertical plane. The sign of $J(□, B)$ can be determined by the similar rule. These simple rules are useful for the mode structure analyses.

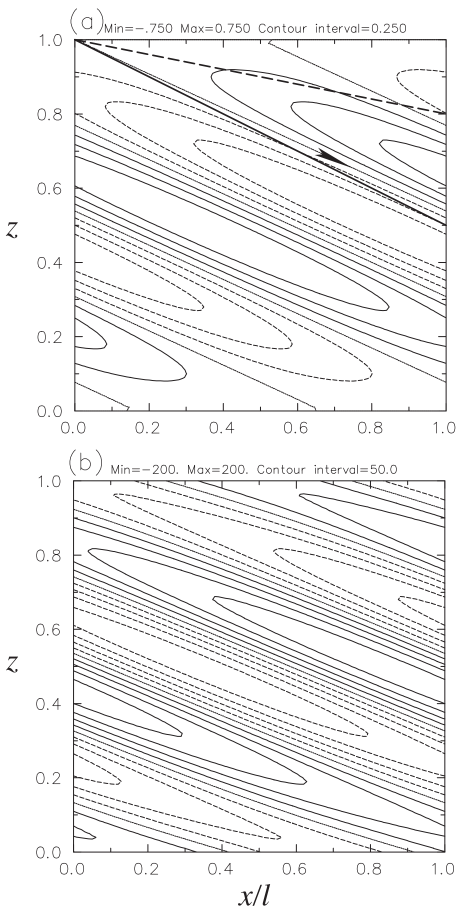

Fig. 2. □ (a) and G (b) plotted (solid for positive, dotted for zero and dashed for negative) in the cross-band vertical plane $(x/l, z)$ for the fastest growing mode with $□_+ = 1.14$ at the parameter point of $Ri = 0.4$ and $l = 0.2$ (case 1). The circulation direction is shown by the arrow in (a). The slopes of M-surface and B-surface are shown in (a) by the dashed and solid lines, respectively.

b. Growing and decaying modes

Figure 2 shows the structures of □ and G for the fastest growing mode (for which n =1) at the parameter point of $Ri = 0.4$ and $l = 0.2$ for case 1 in Fig. 1. The growth rate is $□_+(1) = 1.143$. As in (2.8), to facilitate the comparisons between different modes, the amplitude of □ is set to unity in Fig. 2a, where the dashed and solid lines are M-surface and B-surface intersected by the cross-band vertical plane, respectively. As shown by the arrow along the streamline between the two cells, the slantwise downdraft is steeper than the M-surface but less steep than the B-surface in (x, z). Thus, according to the above simple rules, $J(□, M)$ is positive and $J(□, B)$ is negative along the downdraft. Since these two terms are the only generation terms in (3.1b)-(3.1c), their generated v is positive and b is negative along the downdraft (not shown). The Coriolis force associated with the positive v is rightward and the buoyancy associated with the negative b is downward. Their combined vector force accelerates the slantwise downdraft. The slantwise updraft is accelerated similarly by this type of positive feedback. The overall positive feedback is indicated by the positive correlation between the vorticity and vorticity generation or, equivalently, by the negative correlation between G (Fig. 2b) and □ (Fig. 2a). For the decaying mode at $Ri = 0.4$ and $l = 0.2$, the growth rate is negative and given by $□_+(1) = -1.143$. In this case, the mode structures (not shown) are the same as those in Fig. 2 except that the G field changes sign from that in Fig. 2b and thus the feedback becomes negative.

The above positive (or negative) feedback is seen in general for all the growing (or decaying) modes. According to (2.8a), the streamline slope is given by $dz/dx|_□ = -□^{-1}$ along the middle level for $n = 1$ (or along the levels of $\cos(n□z) = 0$ for $n > 1$). Note that $□^{-1} = 1 + □_+^2 > 1$ for $□_+^2 > 0$. One can verify that $□^{-1} = 1 + □_+^2 < Ri^{-1}$. Combining these two conditions gives

$$-1 > -□^{-1} > -Ri^{-1} \tag{3.3}$$

for any growing or decaying mode. This means that the streamlines are all slantwise backward (with a slope of $-□^{-1}$ at the middle level) between the M-surface (slope of -1) and B-surface (slope of $-Ri^{-1}$) for the growing and decaying modes with n =1. This general feature has been seen from the example in Fig. 2a. The backward slantwise streamlines indicate that u and $-w$ are correlated positively and their wave patterns are exactly in-phase along the middle level. According to (2.8), $-v$ and b are related to $w = ik□$ by $(1 - □)/□_+$ and $(□ - Ri)/□_+$, respectively, along the middle level. These two factors are positive for the growing modes according to (3.3). Thus, the buoyancy b is in-phase with w and the Coriolis force associated with v is in-phase with u for the growing modes. For the decaying modes, the above in-phase relationships change into completely out-of-phase ones and thus the feedback becomes negative.

c. Fast propagating modes

Figure 3 shows the structure of □ for the fastest rightward propagating mode (with n =1) at the same parameter point as that (Ri = 0.4 and l = 0.2 for case 1) in Fig. 2. For the reason explained in section 2d, the propagation mechanism can be examined based on the middle-level phase relationships between the cross-band motions and restoring forces. As shown in Fig. 3, the middle-level streamlines are slantwise forward while the M-surface and B-surface are tilted backward. Thus, J(□, M) and J(□, B) are both positive following the leftward-downward motion (shown by the arrow in Fig. 3). Their generated v and b are positive but lag the leftward-downward motion by a quarter of wavelength as the wave propagates rightward (not shown). The associated Coriolis force and buoyancy are rightward and upward, respectively. The reverse motion produced by this vector force is thus lagged by a half wavelength with respect to the motion that generates this vector force. This explains how the modes propagate. The propagation mechanism is also indicated by the lagged correlation (by a quarter of wavelength) between the vorticity and vorticity generation, or equivalently, between G (not shown) and □.

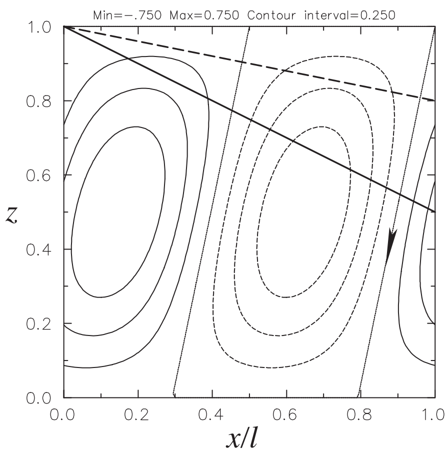

Fig. 3. As in Fig. 2a but for the fastest rightward propagating mode. The phase speed is $-i$□$_-(1)/k$ = 5.0.

The above mechanism holds for all the fast propagating modes associated with □$_-(n)$. Since □$_-^2 < -1$ for all □$_-$ = □$_-(n)$ (with n = 1, 2, ...), we have

$$-\square^{-1} > 0 \qquad (3.4)$$

for any fast propagating modes. This means that the streamlines are tilted forward along the middle level as seen from the example in Fig. 3. The forward slantwise streamlines indicate that u and w are correlated positively and their wave patterns are exactly in-phase along the middle level. According to (2.8), v and b are related to $w = ik$□ by (□ - 1)/□$_-$ and (□ - Ri)/□$_-$, respectively, along the middle level (for n = 1). Here, □$_-$ is purely imaginary, □ - 1 and □ - Ri are both negative according to (3.4). This means that v and b are exactly in phase along the middle level and their wave patterns lead those of u and w by 90^o (a quarter of wavelength in the direction of the mode propagation). Along the middle level, p_x = 0 (not shown), the Coriolis force associated with v is the only horizontal restoring force and its wave pattern leads that of u by 90^o. The buoyancy b is partially offset by the vertical pressure gradient force and thus the middle-level wave pattern of the net vertical restoring force b - p_z leads that of w by 90^o. These phase relationships support the above analysis of the propagation mechanism.

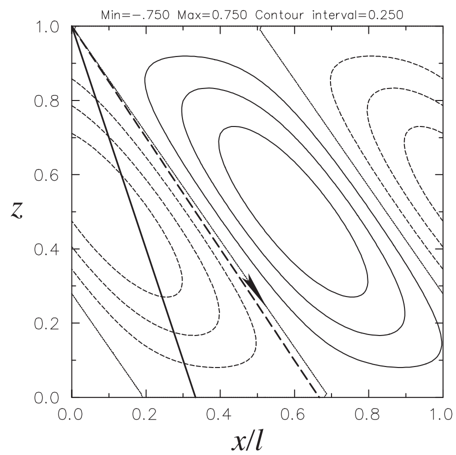

Fig. 4. As in Fig. 3 but for the slowest leftward propagating mode at the parameter point of Ri = 0.5 and l = 1.5 (case 3 in Fig 1). The phase speed is $-i$□$_+(1)/k$ = 0.20.

d. Slowly propagating modes

Figure 4 shows the structure of □ for the slowest leftward propagating mode (with $n = n_s = n_c + 1 = 1$) at the parameter point of Ri = 0.5 and l = 1.5 for case 3 in Fig 1. As shown in Fig. 4, the middle-level streamlines are slantwise slightly more backward than the M-surface and much more backward than B-surface. Thus, J(□, M) is slightly negative and J(□, B) is strongly negative following the downward motion (shown by the arrow in Fig. 4). Their generated v and b are negative and lag the downward motion by a quarter of wavelength as the wave propagates rightward (not shown). The associated Coriolis force is weak but still leftward. The associated buoyancy is downward but the vertical gradient of the perturbation pressure is upward according to (3.2b). The combined vector force, plus the perturbation pressure gradient, is weak but still tends to reverse the motion and thus makes the wave propagating slowly. The propagation mechanism is also indicated by the lagged correlation (by a quarter of wavelength) between G (not shown) and ψ. The structures for the slowest leftward propagating mode (not shown) are the same as those for the rightward propagating mode in Fig. 4 except that the v, b and G fields are shifted by a half of wavelength.

The above mechanism holds for all the slowly propagating modes associated with $\sigma_+(n) < 0$. Note that $0 > \sigma_+^2 > -1$ for $n > n_c$ [see (2.7)], so $1 > \beta^{-1} = 1 + \sigma_+^2 > 0$ or, equivalently,

$$0 > -\beta^{-1} > -1 \tag{3.5}$$

for any slowly propagating modes. This means that the middle-level streamlines are tilted backward and more slantwise than the M-surface and B-surface along the middle level as seen from the example in Fig. 4. The backward slantwise streamlines indicate that u and w are correlated negatively and their middle-level wave patterns are completely out-of-phase. Along the middle level, v and b are related to w by $(\beta -1)/\sigma_+$ and $(\beta - Ri)/\sigma_+$, respectively. Here, σ_+ is purely imaginary, but $\beta - 1$ and $\beta - Ri$ are positive according to (3.5). This means that the middle-level wave pattern of v leads that of u by 90^o but the middle-level wave pattern of b lags that of w by 90^o. Here, the phase relationship between v and u is similar to that for the fast propagating modes and the Coriolis force associated with v is still the only horizontal restoring force along the middle level. The phase relationship between b and w, however, is opposite to that for the fast propagating modes. The buoyancy b is overly offset by the vertical pressure gradient force $-p_z$, so the middle-level wave pattern of the net vertical restoring force $b - p_z$ leads that of w by 90^o. This net vertical restoring force, however, is weaker than that for the fast propagating modes.

Since the horizontal perturbation pressure gradient force vanishes along the middle level (for $n = 1$), the middle-level cross-band horizontal motion and restoring forcing are related to each other quite similarly for the fast and slowly propagating modes. The middle-level vertical motion and restoring forcing, however, are related to each other in different ways for the fast and slowly propagating modes. The vertical motion is positively (or negatively correlated to the cross-band horizontal motion for the fast (or slowly) propagating modes. The perturbation buoyancy is only partially offset by the vertical perturbation pressure gradient force for the fast propagating modes but is overly offset by the vertical perturbation pressure gradient force for the slowly propagating modes. These differences are tied up with the slopes of the middle-level streamlines. For the fast propagating modes, the streamlines are tilted forward in the opposite direction with respect to the M-surface and B-surface according to (3.4). For the slowly propagating modes, the streamlines are tilted backward and more slantwise than the M-surface and B-surface according to (3.5). The above differences suggest that the propagation of the fast propagating modes is driven by both the Coriolis inertial restoring force and perturbation buoyancy restoring force, while the propagation of the slowly propagating modes is driven by the Coriolis inertial restoring force but is slowed by the perturbation buoyancy restoring force (because the perturbation buoyancy tends to drive the wave propagation in the opposite direction). This explains why the slowly propagating modes propagate more slowly than the fast propagating modes (for $n = 1$]. The above analysis can be done similarly for any $n > 1$, because the mode structure in each layer between two adjacent levels of $w = 0$ [$\sin(n\pi z) = 0$] is similar to that for $n = 1$.

4. Partial orthogonality

The normal modes obtained in (2.8) can be conveniently numbered by $j = 2(n - 1)\mathrm{sgn}(m) + m$, where n (= 1, 2, ...) is the vertical-mode number n (= 1, 2, ...) and m (= ±1, ±2) is the root number for the four roots ($\pm\sigma_+$, $\pm\sigma_-$) of (2.5). Here, $j = j(n, m)$ can be viewed as an integer function of n and m. According to this and (2.8a), the j-th ψ-component mode is then given by $\psi_j \exp(\sigma_j t)$, where $\psi_j = \exp[-ik(x + \beta_j z)]\sin(n\pi z)$ is the mode structure at the initial time ($t = 0$) with $\sigma_j = \sigma(n, m)$ and $\beta_j = (1 + \sigma_j^2)^{-1}$. Substituting $\psi_j \exp(\sigma_j t)$ back into (2.2) gives

$$(\sigma_j^2 \Delta_a + P)\psi_j = 0, \tag{4.1a}$$
$$\psi_j = 0 \quad \text{at } z = 0, 1. \tag{4.1b}$$

where $P(\ ) = Ri\,(\ )_{xx} - 2(\ )_{xz} + (\ )_{zz}$. Note that Δ_a and P are self-adjoint and σ_j^2 is real, so the eigenvalue problem in (4.1) is self-adjoint and a partial-orthogonality condition can be derived for ψ_j (j = ±1, ±2, ...) below.

Denote by ψ_j^* the complex conjugate of ψ_j, so the complex conjugate of (4.1) is

$$(\sigma_j^2 \Delta_a + P)\psi_j^* = 0, \tag{4.2a}$$
$$\psi_j^* = 0 \quad \text{at } z = 0, 1. \tag{4.2b}$$

By averaging $\psi_{j'}$(4.2a) over one-wavelength area in the cross-band vertical section and using integration by parts with the boundary conditions (4.1b) and (4.2b) and periodic conditions in the horizontal, we obtain

$$\sigma_j^2\{\langle(\nabla_a\psi_j)^H(\nabla_a\psi_{j'})\rangle\} = \{\langle\psi_{j'}P\psi_j^*\rangle\}$$
$$= \{\langle\psi_j^* P\psi_{j'}\rangle\}, \tag{4.3a}$$

where $\{(\ )\}$ denotes the vertical average and $\langle(\ )\rangle$ denotes the horizontal average of ( ) in the cross-band vertical section over one wavelength, $\nabla_a = (a(\ )_x, (\ )_z)^T$ denotes the scaled gradient operator, $(\ )^H = (\ )^{*T}$ the Hermit transpose of the vector ( ), and $(\ )^T$ the transpose of ( ). Similarly, by setting the mode number to j' (in place of j) in (4.1a) and averaging ψ_j^*(4.1a), we obtain

$$\sigma_{j'}^2\{<(\nabla_a\psi_j)^H(\nabla_a\psi_{j'})>\} = \{<\psi_j^* P\psi_{j'}>\}. \qquad (4.3b)$$

The difference between (4.3a) and (4.3b) yields

$$(\sigma_j^2 - \sigma_{j'}^2)[\psi_j, \psi_{j'}]_{K2} = 0, \qquad (4.4a)$$

where $[\psi_j, \psi_{j'}]_{K2} = \{<(\nabla_a\psi_j)^H(\nabla_a\psi_{j'})>\} = \{<(u_j^* u_{j'} + a^2 w_j^* w_{j'}>\}$ is the inner product associated with the cross-band kinetic energy defined by

$$\{K_2\} = \{<|u|^2 + a^2|w|^2>/2\}. \qquad (4.4b)$$

Note that $\sigma_j^2 - \sigma_{j'}^2 \neq 0$ unless $|j| = |j'|$, so (4.4a) leads to the following orthogonality:

$$[\psi_j, \psi_{j'}]_{K2} = 0 \quad \text{for} \quad |j| \neq |j'|. \qquad (4.5a)$$

This result can be also derived by substituting the analytical expressions of ψ_j^* and $\psi_{j'}$ into $[\psi_j, \psi_{j'}]$. One can verify that

$$[\psi_j, \psi_{j'}]_{K2} = p_j = [(n\pi)^2 + (a^2 + \beta_j^2)k^2]/2 \quad \text{for} \quad |j| = |j'|. \qquad (4.5b)$$

As shown in (4.5), measured by the inner product defined in (4.4), ψ_j and $\psi_{j'}$ are orthogonal between different pairs but parallel within each pair. Because of this and the reason explained below, the orthogonality is considered to be partial. Note that $[\psi_j\exp(\sigma_j t), \psi_{j'}\exp(\sigma_{j'} t)]_{K2} = \exp(\sigma_j^* t + \sigma_{j'} t)[\psi_j, \psi_{j'}]_{K2}$, so the orthogonality in (4.5a) is equally applicable for the ψ-component modes at any given time (not limited to the initial time).

Note from $j = 2(n - 1)\mathrm{sgn}(m) + m$ that $j = -j'$ is equivalent to $n = n'$ and $m = -m'$, so ψ_j and ψ_{-j} are identical. This means that the two ψ-component modes $\psi_j\exp(\sigma_j t)$ and $\psi_{-j}\exp(\sigma_{-j} t)$ are identical at the initial time and thus cannot be separated from each other in the initial field of ψ without utilizing the time tendency information provided by ψ_t which is contained implicitly in the initial conditions for (2.2). Because ψ_j and ψ_{-j} are identical, the streamfunction space is spanned by ψ_j with j going through either all positive or all negative integers but not both. This space is complete for ψ-component fields (of wavelength l at any given time) but is incomplete for the full-component fields in the initial conditions, such as (ψ, v, b) in (2.2). To utilize all the information in the model initial conditions, it is necessary to expand the streamfunction space to a vector-function space spanned by $(\psi_j, \sigma_j\psi_j)^T$ or $(\psi_j, v_j, b_j)^T$ with j going through all positive and negative integers. Here, (v_j, b_j) denotes the initial fields of the j-th (v, b)-component modes in (2.8b,c). In such an expanded vector-function space, the inner product defined in (4.4) is no longer a full metric. However, a full metric can be given by the total perturbation energy defined by

$$E = \{K_2\} + \{K_v + P_b\}, \qquad (4.6)$$

where $K_v = <|v|^2>/2$ the kinetic energy associated with the along-band velocity perturbation (called along-band kinetic energy) and $P_b = <|b|^2>Ri/2$ is the potential energy associated with the buoyancy perturbation (called buoyancy energy). With this metric, the full-component modes in (2.8) are non-orthogonal (at any given time). This further explains why the orthogonality in (4.5) is considered to be partial.

5. Non-modal solutions and singular vectors

Consider that the non-modal solutions are periodic with a given wavenumber k in the horizontal, so the non-modal solution can be constructed by

$$\psi(x, z, t) = \sum_j c_j\psi_j\exp(\sigma_j t) = \mathbf{c}^H\boldsymbol{\psi}, \qquad (5.1)$$

where $\psi_j\exp(\sigma_j t) = \exp[\sigma_j t - ik(x + \beta_j z)]\sin(n\pi z)$ is the j-th streamfunction mode as in (4.1), $\beta_j = (1 + \sigma_j^2)^{-1}$, c_j is a complex coefficient for the j-th mode, the summation $\sum_j$ is over j (= ±1, ±2, ...), $\mathbf{c}$ is the vector composed of c_j, $\mathbf{c}^H$ is the Hermitian transpose of $\mathbf{c}$, and $\boldsymbol{\psi}$ is the vector composed of $\psi_j\exp(\sigma_j t)$. As in section 4, the mode is numbered by $j = 2(n - 1)\mathrm{sgn}(m) + m$,, so $j = -j'$ is equivalent to $n = n'$ and $m = -m'$. This implies that $\sigma_{-j} = -\sigma_j$, $\beta_{-j} = \beta_j$, and $\psi_{-j} = \psi_j$.

Substituting (5.1) into $(u, w) = (\psi_z, -\psi_x)$ and (2.2b)-(2.2c) gives

$$(u, w, v, b) = \mathbf{c}^H(\mathbf{u}, \mathbf{w}, \mathbf{v}, \mathbf{b}) \qquad (5.2)$$

where ($\mathbf{u}$, $\mathbf{w}$, $\mathbf{v}$, $\mathbf{b}$) are the vectors composed of $(u_j, w_j, v_j, b_j)\exp(\sigma_j t)$, $u_j = \psi_{jz}$, $w_j = -\psi_{jx}$, $v_j = (\psi_{jx} - \psi_{jz})/\sigma_j$, and $b_j = (\psi_{jx} - Ri^{-1}\psi_{jz})/\sigma_j$. Substituting (5.2) into the squared norm defined by the total perturbation energy in (4.6) gives

$$E = E(t) = \mathbf{c}^H\mathbf{A}(t)\mathbf{c}, \qquad (5.3)$$

where $\mathbf{A}(t) = \{<\mathbf{u}\mathbf{u}^H + a^2\mathbf{w}\mathbf{w}^H + \mathbf{v}\mathbf{v}^H + \mathbf{b}\mathbf{b}^H Ri>\}/2$ is a matrix function of t. Measured by the inner product associated with the total perturbation energy norm, the normal modes in (5.2) are not orthogonal, so $\mathbf{A}(t)$ contains non-diagonal terms. This implies that the non-modal energy growths can be larger than the modal growths. The non-modal energy growth from $t = 0$ to a specified optimization time $t = \tau$ is measured by

$$\lambda = \lambda(\tau) = E(\tau)/E(0) = \mathbf{c}^H\mathbf{A}(\tau)\mathbf{c}[\mathbf{c}^H\mathbf{A}(0)\mathbf{c}]^{-1}. \qquad (5.4)$$

The energy growth is maximized when **c** is the eigenvector associated with the largest eigenvalue of the following eigenvalue problem:

$$[\mathbf{A}(\square) - \square\mathbf{A}(0)]\mathbf{c} = 0. \qquad (5.5)$$

The largest eigenvalue denoted by $\square_{max}$ is called the leading singular value, while the associated eigenvector is called the leading singular vector and is denoted by $\mathbf{c}_{ls}$. The solution given by $\square = \mathbf{c}_{ls}^{H}\square$ or $(u, w, v, b) = \mathbf{c}_{ls}^{H}(\mathbf{u}, \mathbf{w}, \mathbf{v}, \mathbf{b})$ is called the leading singular perturbation that has the maximum energy growth at the optimization time $t = \square$. According to the numerical solutions of (5.5) (not shown), the maximum non-modal growth is often produced dominantly by two paired modes. Non-modal growths produced by paired modes are examined analytically in the next section.

6. Non-modal growths produced by paired modes

a. Paired propagating modes

Consider a pair of propagating modes, say, the j-th pair composed of the j-th and j'-th modes with $j = -j' > 0$. According to (2.8), these two modes have the same, exactly in-phase, spatial structures in (u, w) but the opposite, exactly 180° out-of-phase, spatial structures in (v, b). These two modes propagate in opposite horizontal directions and their phase speeds are given by $\square_j/k$ (> 0) and $\square_{j'}/k = -\square_j/k$ (< 0), respectively, where $\square_j = \square_j/i$ and $\square_{j'} = \square_{j'}/i$ are their respective frequencies. Denote by $\mathbf{A}_j(t)$ the 2x2 submatrix of $\mathbf{A}(t)$ associated with the j-th subspace spanned by the j-th and j'-th modes with $j = -j' > 0$. By using the analytical form of the normal-mode solution in (2.8), one can show that $\mathbf{A}_j(t)$ has the following form:

$$\mathbf{A}_j(t) = X_j \begin{bmatrix} 1 & \exp(i2\square_j t) \\ \exp(\square i2\square_j t) & 1 \end{bmatrix} + Y_j \begin{bmatrix} 1 & \square\exp(i2\square_j t) \\ \square\exp(\square i2\square_j t) & 1 \end{bmatrix}, \qquad (6.1)$$

where

$$X_j = \{<|u_j|^2 + a^2|w_j|^2>\}/2 = [(n\square)^2 + (a^2 + \square_j^2)k^2]/4, \qquad (6.2a)$$

$$Y_j = \{<|v_j|^2> + Ri<|b_j|^2>\}/2 = Z_j/|\square_j|^2, \qquad (6.2b)$$

$$Z_j = [X_j + (Ri - a^2)k^2/4](1 + Ri^{-1}) - \square_j k^2. \qquad (6.2c)$$

Here, $\square_j = \square_j/i = -\square_{j'}/i = -\square_{j'} > 0$, $\square_j = \square_{j'}$ and the aforementioned opposite polarization relationships between the two paired propagating modes are used in the derivation of (6.1)-(6.2). According to (4.6), X_j is the cross-band kinetic energy and Y_j is the along-band kinetic energy plus the buoyancy energy for the j-th mode at the initial time. In the j-th subspace, the eigenvalue problem in (5.5) reduces to

$$[\mathbf{A}_j(\square) - \square_j\mathbf{A}_j(0)]\mathbf{c}_j = 0, \qquad (6.3)$$

where $\mathbf{c}_j = (c_j, c_{j'})^T = (c_j, c_{-j})^T$ is the vector coefficient for the j-th and j'-th modes and $(\ )^T$ denotes the transpose of ( ). Here, $\square_j$ denotes the eigenvalue in the j-th subspace. Since the solution will be considered only the subspace, the subscript j will be dropped from $\square_j$ as long as the meaning is clearly understood.

One can verify that (6.3) has two eigenvalues given by

$$\square_\pm = q_j \pm (q_j^2 - 1)^{1/2}, \qquad (6.4)$$

where

$$q_j = [1 - \square^2\cos(2\square_j\square)]/(1 - \square^2) \qquad (6.5)$$

and

$$\square = (X_j - Y_j)/(X_j + Y_j). \qquad (6.6)$$

Here, q_j and $\square_\pm$ are periodic functions of $\square$. The function forms of $\square_+$ are plotted in Fig. 5 (for $\square^2$ = 0, 0.2, 0.4, 0.6, 0.8) over one period ($0 \square 2\square_j\square \square 2\square$). When $\square$ increases from 0 to $\square/(2\square_j)$, $\square_+$ increases from 1 to $\max(X_j/Y_j, Y_j/X_j)$ and $\square_\square$ decreases from 1 to $\min(X_j/Y_j, Y_j/X_j)$. Thus, as long as $X_j \square Y_j$, a non-modal energy growth can be caused by the paired propagating modes and the maximum growth is $\square_+ = \max(X_j/Y_j, Y_j/X_j)$ as $\square = \square/(2\square_j)$. The associated eigenvector is $\mathbf{c}_j = (c_j, c_{j'})^T \square (1, -1)^T$ if $X_j > Y_j$ or $\mathbf{c}_j = (c_j, c_{j'})^T \square (1, 1)^T$ if $X_j < Y_j$. For the paired fastest propagating modes, $\square_+ = Y_j/X_j > 1$ always. For the paired slowest modes, $\square_+ = = Y_j/X_j = 1$ occurs only when the parameter point is at the zero $\square$ contour (not shown but in the region of $Ri > 1$).

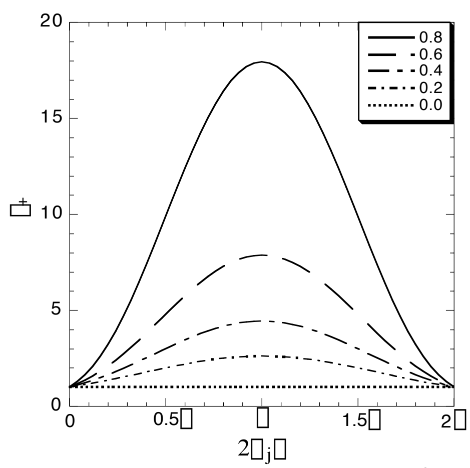

Fig. 5. $\square_+$ plotted for different values of $\square^2$ (= 0, 0.2, 0.4, 0.6, 0.8) over one period ($0 \square 2\square_j\square \square 2\square$). Here, $\square_+$ is the leading singular value in the subspace spanned by a pair of propagating modes.

As explained at the beginning of this section, the two paired propagating modes have exactly opposite polarization relationships between ψ and (v, b). As the two modes propagate toward each other in opposite horizontal directions, their associated ψ and (u, w) fields become exactly in-phase (or out-of-phase) when their two associated (v, b) fields become exactly out-of-phase (or in-phase). Thus, the composed (v, b) fields oscillate with the same frequency ω_j as the composed (u, w) fields but the oscillations of the (v, b) fields are lagged by 90^o with respect to the oscillations of the (u, w) fields. Note that $\{K_2\}$ and $\{K_v + P_b\}$ are integrated squares of (u, w) and (v, b), respectively, so they oscillate between 0 and their respective maxima, with the same frequency of $2\omega_j$ and the phase difference between the oscillations of $\{K_2\}$ and $\{K_v + P_b\}$ is just 180^o. Since the amplitudes of the composed (u, w) and (v, b) are twice of those for the j-th or j'-th mode, the maxima of $\{K_2\}$ and $\{K_v + P_b\}$ are $4X_j$ and $4Y_j$, respectively. When $X_j = Y_j$, the oscillation of $\{K_2\}$ offsets the oscillation of $\{K_v + P_b\}$, so the total energy $E = \{K_2\} + \{K_v + P_b\}$ keeps constant in time. This explains why the paired propagating modes produce no energy growth (that is, $\lambda_+ = \lambda_- = 1$) when $X_j = Y_j$.

However, when $X_j > Y_j$ (or $X_j < Y_j$), the oscillation of $\{K_2\}$ (or $\{K_v + P_b\}$) becomes dominant and thus the total energy oscillates between $4Y_j \le E \le 4X_j$, (or $4X_j \le E \le 4Y_j$). In particular, if $X_j > Y_j$ and $\mathbf{c}_j \propto (1, -1)^T$, then $\{K_2\}= 0$ and thus $E = \{K_v + P_b\} = 4Y_j$ at $t = 0$. As t increases from 0 to $\pi/2\omega_j$ (that is, one quarter of the wave period of the j-th mode), $\{K_2\}$ increases from 0 to the maximum (= $4X_j$) but $\{K_v + P_b\}$ decreases from the maximum (= $4Y_j$) to 0, so E increases maximally from $4Y_j$ to $4X_j$. Similarly, if $X_j < Y_j$ and $\mathbf{c}_j \propto (1, 1)^T$, then E increases maximally from $4X_j$ to $4Y_j$ as t increases from 0 to $\pi/2\omega_j$. This explains why and how the paired propagating modes produce the energy growth of $\lambda_+ = X_j/Y_j$ for $X_j > Y_j$ (or $\lambda_+ = Y_j/X_j$ for $X_j < Y_j$) as t increases from 0 to $\tau = \pi/2\omega_j$. Note from (6.2) that Y_j/X_j is proportional to $|\sigma_j|^{-2} = \omega_j^{-2}$, so λ_+ can be very large when ω_j is very small.

b. Paired growing and decaying modes

Consider a pair of growing and decaying modes, say, the j-th pair. Denote by $\mathbf{A}_j(t)$ the 2x2 submatrix of $\mathbf{A}(t)$ associated with the j-th subspace spanned by the j-th pair of modes (with $j = -j' > 0$). By using (2.8) and considering the aforementioned polarization relationships, one can show that

$$\mathbf{A}_j(t) = X_j\begin{pmatrix} \exp(2\sigma_j t) & 1 \\ 1 & \exp(-2\sigma_j t)\end{pmatrix}$$

$$+\ Y_j\begin{pmatrix} \exp(2\sigma_j t) & -1 \\ -1 & \exp(-2\sigma_j t)\end{pmatrix}, \tag{6.7}$$

where X_j and Y_j are as in (6.2) but $|\omega_j| = |\sigma_j| = \sigma_j = -\sigma_{j'} > 0$. In the j-th subspace, the eigenvalue problem in (5.5) reduces to the same form as that in (6.3) but $\mathbf{A}_j(\tau)$ and $\mathbf{A}_j(0)$ are given by (6.7) instead of (6.1). The reduced eigenvalue problem with (6.7) has two eigenvalues given by $\lambda_\pm = q_j \pm (q_j^2 - 1)^{1/2}$ as in (6.4) but with

$$q_j = [\cosh(2\sigma_j\tau) - \gamma^2]/(1 - \gamma^2), \tag{6.8}$$

where γ is as in (6.6). If $X_j = Y_j$, then $\gamma = 0$ and the two eigenvalues reduce to $\lambda_\pm = \exp(\pm 2\sigma_j\tau)$. In this case, the energy growth is supported solely by the growing mode and thus is not affected by the decaying mode. This occurs only when the parameter point is on the zero γ-contour line inside the unstable region (not shown). Away from the zero γ-contour line, we have $X_j \ne Y_j$ and $(1 - \gamma^2)^{-1} = 1 + (X_j - Y_j)^2/(4X_jY_j) > 1$ within the unstable region. The scaled non-modal growth, $\lambda_+\exp(-2\sigma_j\tau)$, increases from 1 to the asymptotic limit, $(1 - \gamma^2)^{-1}$, as τ increases from 0 to infinity.

The eigenvector associated with λ_+ is given by $\mathbf{c}_j = (c_j, c_{j'})^T \propto (1, -\gamma)^T \propto (X_j + Y_j, Y_j - X_j)^T$. By setting $c_j = X_j + Y_j$ and $c_{j'} = Y_j - X_j$, the initial non-modal fields are given by $c_j(u_j, w_j) + c_{j'}(u_{j'}, w_{j'}) = 2Y_j(u_j, w_j)$ and $c_j(v_j, b_j) + c_{j'}(v_{j'}, b_{j'}) = 2X_j(v_j, b_j)$, so the initial total energy is $E(0) = \{K_2\} + \{K_v + P_b\} = (2Y_j)^2X_j + (2X_j)^2Y_j = 4X_jY_j(X_j + Y_j)$ according to (5.3), (6.2) and (6.7). The initial total energy for the growing mode only, however, is $c_j^2(X_j + Y_j) = (X_j + Y_j)^2(X_j + Y_j)$. Thus, the total energy E is reduced at the initial time by a factor of $4X_jY_j(X_j + Y_j)^{-2} = 1 - \gamma^2$ (< 1) due to the inclusion of the decaying mode. Thus, as $t \to \tau \to \infty$, the decaying mode diminishes and $\exp(-2\sigma_j\tau)E(\tau)/E(0) \to (X_j + Y_j)^2(4X_jY_j)^{-1} = (1 - \gamma^2)^{-1}$. This is precisely the above derived asymptotic limit of the scaled non-modal growth.

In the above analysis, $c_j = X_j + Y_j$ is the coefficient for the growing mode and is always positive, while $c_{j'} = Y_j - X_j$.is the coefficient for the decaying mode. Since these two paired modes have opposite polarization relationships between (u, w) and (v, b), we have $(u_j, w_j) = (u_{j'}, w_{j'})$ and $(v_j, b_j) = -(v_{j'}, b_{j'})$. When $X_j > Y_j$, $c_{j'}$ is negative, the initial fields $c_{j'}(u_{j'}, w_{j'})$ for the decaying mode and $c_j(u_j, w_j)$ for the growing mode are completely out of phase. This causes an decrease in $\{K_2\}$ that overly offsets the increase in $\{K_v + P_b\}$ caused by the in-phase relationship between $c_j(v_j, b_j)$ and $c_{j'}(v_{j'}, b_{j'})$ at the initial time. When $X_j < Y_j$, $c_{j'}$ is positive, so the initial fields $c_j(v_j, b_j)$ and $c_{j'}(v_{j'}, b_{j'})$ are completely out off phase. This causes an decrease in

$\{K_v + P_b\}$ that overly offsets the increase in $\{K_2\}$ caused by the in-phase relationship between $c_{j'}(u_{j'}, w_{j'})$ and $c_j(u_j, w_j)$ at the initial time. Thus, as long as $X_j \neq Y_j$, the decaying mode can reduce the initial total energy and enhance the non-modal growth.

c. Paired stationary modes

As shown in section 2c, when $Ri \to Ri_{nc} = 1 - (n_c l/2)^2$, $\sigma_+(n_c) \to 0$ and the associated pair of modes degenerated into a pair of stationary modes. For such a pair of stationary modes, we have $j = 2(n - 1)\mathrm{sgn}(m) + m = \pm(2n_c + 1)$. Denote by $\mathbf{A}_j(t)$ the 2x2 submatrix of $\mathbf{A}(t)$ associated with the j-th subspace spanned by this pair of stationary modes. By using the analytical forms of the two stationary modes in (2.9)-(2.10), one can show that $\mathbf{A}_j(t)$ has the following form:

$$\mathbf{A}_j(t) = X_j \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + Z_j \begin{pmatrix} 1 & t \\ t & t^2 \end{pmatrix}, \tag{6.9}$$

where X_j and Z_j are defined as in (6.2) but with $Ri = Ri_{nc}$, $n = n_c$ and $\beta_j^2 = 1$ since $|\omega_j| = |\sigma_j| = 0$. In this case, the eigenvalue problem in (5.5) reduces to the same form as that in (6.3) but $\mathbf{A}_j(\tau)$ and $\mathbf{A}_j(0)$ are given by (6.9). The reduced eigenvalue problem with (6.9) has two eigenvalues given by $\lambda_\pm = q_j \pm (q_j^2 - 1)^{1/2}$ as in (6.4) but with

$$q_j = 1 + \rho^2\tau^2/2, \tag{6.10}$$

where $\rho^2 = Z_j/X_j$. Note that $\lambda_+ \to \rho^2\tau^2 \to \infty$ as $\tau \to \infty$. In this limit, the energy growth λ_+ (in the subspace spanned by the paired stationary modes) is produced entirely by the second stationary mode [see (2.10)].

The result in (6.10) can be also derived from (6.5) or (6.8) in the limit of $|\omega_j| = |\sigma_j| \to 0$. Note that $\gamma = (X_j - Y_j)/(X_j + Y_j) \to -1 + 2|\sigma_j|^2 X_j/Z_j + O(|\sigma_j|^4)$ as $|\sigma_j| \to 0$. Using this result and the Taylor expansion of $\cos(2\omega_j\tau)$ with $\omega_j^2 = -\sigma_j^2$ (> 0), one can verify that (6.5) degenerates into (6.10) in the limit of $|\omega_j| \to 0$. It is also easy to verify that (6.8) degenerates into (6.10) in the limit of $|\sigma_j| \to 0$.

7. Non-modal growth classification

By using (6.4)-(6.6), (6.8) and (6.10), λ_+ can be precisely obtained in any subspace spanned by any paired modes in the complete set of normal modes obtained in section 2. Denote by λ_{j+} the maximum non-modal growth in the j-th subspace (spanned by the j-th paired modes). Denote by $\max(\lambda_{j+})_N = \max\{\lambda_{j+}| j = 1, 2, \ldots 2N\}$ the maximum among all λ_{j+} for $j = 1, 2, \ldots 2N$. The maximum non-modal growth is scaled by $\exp(2\mathrm{Re}\sigma_1\tau)$ and plotted in Fig. 6a for $\tau = 0.5$. Here, $\mathrm{Re}\sigma_1$ is the real part of $\sigma_1 = \sigma_+(1)$, so $\mathrm{Re}\sigma_1 = \sigma_{max}$ in the unstable region and $\mathrm{Re}\sigma_1 = 0$ in the stable region.

The non-modal growth in Fig. 6a has nearly the same pattern as the numerical result (not shown) obtained from (5.5) in the truncated space with $n \le N = 15$, especially over the broad region of $l > 0.5$. This means that the maximum non-modal growth is produced dominantly by the paired fastest propagating modes (with $j = 2$). When l is smaller than 0.5 and decreases continuously (toward zero), $\max(\lambda_{j+})_N$ is given by λ_{4+}, λ_{6+}, λ_{8+}, ... consecutively. For $1.2 > Ri > 0.8$, the maximum non-modal growth is produced ma inly by the j-th paired fast propagating modes with $j = 4, 6, 8, \ldots$ consecutively as l decreases (from 0.5 to 0.1). In the upper-left corner region ($Ri > 1.2$ and $l < 0.5$) and lower-left ($Ri < 0.8$ and $l < 0.5$) corner region, the non-modal growth in Fig. 6a is significantly smaller than the numerical result (not shown).

When the optimization time is increased from $\tau = 0.5$ to 1.0, the scaled non-modal growth is increased significantly in two regions, as shown by Fig. 6b in comparison with Fig. 6a. One region is in the vicinity of the curved boundary of the unstable region below Ri =1, while the other region is marked by the semi-circle counter (of 2.0) centered at $l = 0.1$ and Ri =1.2. In the upper part (1> Ri > 0.7) of the curved region, $\max(\lambda_{j+})_N$ is given by λ_{4+}, λ_{6+}, λ_{8+}, ... as l becomes smaller than 2.0, 1.0, 0.5, ..., respectively. In the lower part ($Ri < 0.7$) of the curved region, $\max(\lambda_{j+})_N$ is given by λ_{1+}. Note that λ_{1+} is the maximum non-modal growth produced by the paired slowest propagating modes (or by the paired fastest growing and decaying modes) when the parameter point (Ri, l) is outside (or inside) the unstable region.

When the optimization time is increased further to $\tau = 5.0$, the scaled non-modal growth is increased sharply in the banana-shaped region along the boundary of the unstable region (see Fig. 6c). The scaled non-modal growth is also increased by 3 times in the semi-circle region, while the center of the semi-circle region is shifted slightly down to Ri =1.1 (Fig. 6c). In these two regions and outside the unstable region, the scaled growth is very close to that (not shown) computed in the truncated space (with $n \le N = 15$). In this case, $\max(\lambda_{j+})_N$ is given by λ_{1+}, so the maximum non-modal growth is produced almost solely by the paired slowest propagating modes. Inside the unstable region, $\max(\lambda_{j+})_N$ is also given by λ_{1+}, but λ_{1+} is produced by the paired fastest growing and decaying modes.

Note that the semi-circle region (marked by the contour of 2.5) in Fig. 6c largely coincides with the semi-circle region of $\gamma > 0.8$ centered at $l = 0.1$ and Ri =1.1 (not shown). In this region, we have $\gamma > 0.7$ and thus $X_j/Y_j = (1 + \gamma)/(1 - \gamma) > 9.0$ for j = ±1 according to (5.6). In this case, since $X_j > Y_j$, the non-modal growth of the total perturbation energy (produced by the paired slowest propagating modes with j = ±1) is characterized by the increase of the cross-band circulation kinetic energy $\{K_2\}$ that overly offsets the decrease of $\{K_v + P_b\}$. In particular, as shown in section 6a, this type of non-modal growth reaches the maximum of $□_+ = X_j/Y_j$, as $□ = □/(2□_j)$. In this case, $\{K_2\}$ increases from 0 to $4X_j$ and $\{K_v + P_b\}$ decreases from $4Y_j$ to 0 as t increases from 0 to $□ = □/(2□_j)$. This type of non-modal growth requires $X_j > Y_j$ and is classified as PP1 for paired propagating modes.

The banana-shaped region in Fig. 6c largely coincides with the region of $□ < -0.7$ (not shown). In this region, we have $-1 □ □ < -0.8$ and thus $0 □ Y_j/X_j = (1 + □)/(1 - □) < 0.11$ for $j = \pm1$. Here, $□ = -1$ corresponds to $Y_j/X_j = 0$ for $j = \pm1$ while the latter corresponds to $□_1 = 0$ for parameter points along the boundary of the unstable region (see Fig. 1) or, equivalently, along the ridge of the banana-shaped region in Fig. 6c. Immediately outside the unstable region on the long-wavelength side from the ridge of the banana-shaped region in Fig. 6c, the non-modal growth of the total perturbation energy (produced by the paired slowest propagating modes with $j = \pm1$) is characterized by the increase of $\{K_v + P_b\}$ that overly offsets the decrease of $\{K_2\}$. This type of non-modal growth requires $X_j < Y_j$ and is classified as PP2 for paired propagating modes. The PP2 non-modal growth reaches the maximum of $□_+ = Y_j/X_j$, as $□ = □/(2□_j)$. Clearly, the physical mechanism for the PP2 non-modal growth is opposite to that for PP1, although both types of growths are produced by paired propagating modes.

Immediately inside the unstable region on the short-wavelength side from the ridge of the banana-shaped region in Fig. 6c, the non-modal growth (produced by the paired fastest growing and decaying modes) is much larger than the fastest modal growth and the scaled non-modal growth is much larger than one. In this region, $-1 < □ < -0.8$ and $\;\; > (1 - □^2)^{-1} > 2.7$, so the scaled non-modal growth can be very large and very close to its asymptotic limit $(1 - □^2)^{-1}$ as □ is sufficiently large. In this case, as explained in section 4b, since $□ < 0$ and thus $X_j < Y_j$, the non-modal growth is caused by the reduction of $\{K_v + P_b\}$ that overly offsets the increase in $\{K_2\}$ at the initial time due to the inclusion of the decaying mode. This type of non-modal growth requires $X_j < Y_j$ and is classified as GD2 for paired growing and decaying modes.

There is another semi-circle region in Fig. 6c that largely coincides with the semi-circle region of $□ > 0.8$ centered at $l = 0.1$ and $Ri = 0.9$ (not shown). In this region, the scaled non-modal growth has a local maximum at $l = 0.1$ and $Ri = 0.9$ but this maximum is below 2.5 and thus is not shown by the contours (every 2.5) in Fig. 6c. Since $□ > 0$ and thus $X_j > Y_j$ in this region, the non-modal growth is caused by the reduction of $\{K_2\}$ that overly offsets the increase in $\{K_v + P_b\}$ at the initial time due to the inclusion of the decaying mode. This type of non-modal growth requires $X_j > Y_j$ and is classified as GD1 for paired growing and decaying modes.

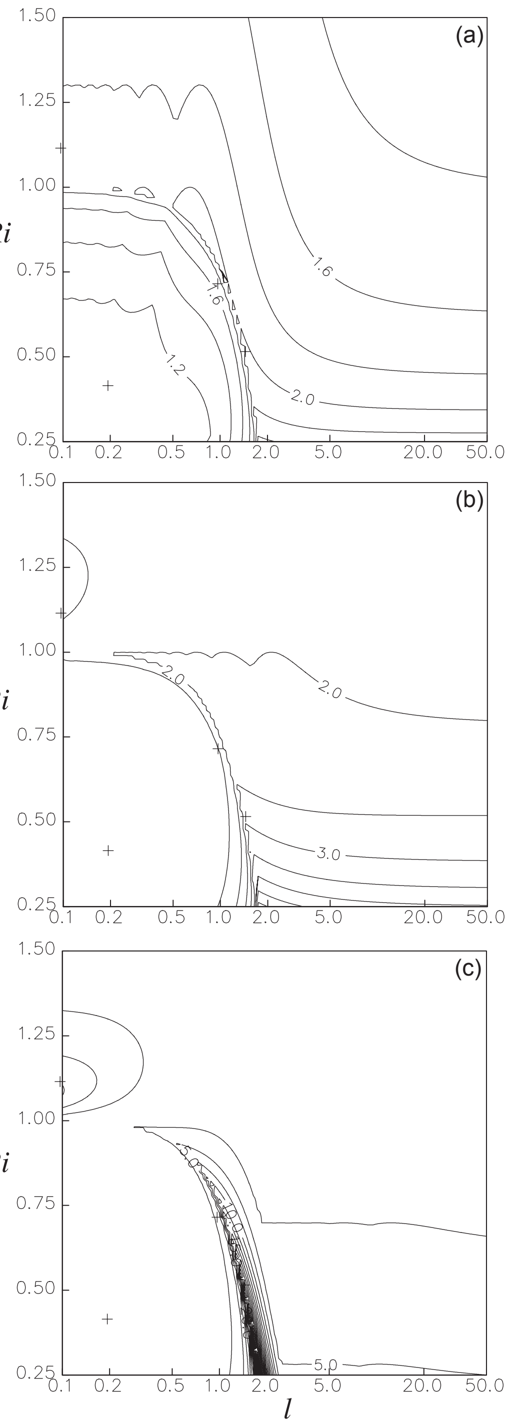

Fig. 6. Contours of $\max(□_{j+})_N \exp(-2\mathrm{Re}□_1□)$ with $N = 15$ plotted in the parameter space of (l, Ri) for $□ = 0.5$ (a), 1.0 (b), and 5.0 (c). The contour intervals are 0.2 in (a), 0.5 in (b), and 2.5 in (c).

8. Conclusions

In this paper, a complete set of normal modes is derived for symmetric perturbations in a vertically sheared basic flow between two horizontal boundaries. The modes can be classified into three types: paired growing and decaying modes, paired slowly propagating modes, and paired fast propagating modes. By rewriting the model system into a two-equation system for the streamfunction ψ and vorticity generation G caused by the thermal-wind imbalance, the interaction between cross-band circulation and its induced perturbation (Coriolis and buoyancy) forces [see (3.1)] becomes intuitive for each type of mode. In particular, for a growing (or decaying mode), the circulation is tilted between the *M*-surface and *B*-surface, so the growth (or decay) of the mode is caused by the positive (negative) feedback between ψ and G. For a slowly propagating mode, the circulation is tilted more slantwise than the M-surface and B-surface, so the mode propagation is driven by the inertial restoring force but slowed by the buoyancy restoring force (because the perturbation buoyancy tends to drive the wave propagation in the opposite direction). For a fast propagating mode, the circulation is tilted in the opposite direction with respect to the M-surface and B-surface, so the mode propagation is driven by both the inertial and buoyancy restoring forces.

The cross-band streamfunction component modes are shown to be orthogonal between different pairs in the scalar-function space in which the orthogonality is measured by the inner product associated with the cross-band kinetic energy [see (4.4)]. The streamfunction alone, however, is insufficient to provide a complete description of the model initial state unless its time tendency information is also utilized. The required time tendency is imbedded in the polarization relationship [see (2.8)]. To utilize the time tendency information, it is necessary to expand the streamfunction space to a vector-function space. The expanded vector-function space is complete for all admissible initial fields. In this space, the full-component normal modes provide a complete set of basis functions and thus can be used to construct any solutions for the initial-boundary value problem governed by the model system.

Measured by the inner-product associated with the total perturbation energy, the full-component modes are non-orthogonal. This implies that the non-modal energy growth produced by a certain linear combination of the normal modes can be larger than the maximum modal growth over a finite time period. Any two paired modes have exactly the opposite polarization relationships. Their cross-band streamfunction component modes are initially identical and thus parallel in the same direction in the associated subspace, and their along-band velocity and buoyancy component modes are also initially parallel but in the opposite directions in the associated subspace. This implies that large non-modal energy growths can be produced by paired normal modes.

The non-modal growths produced by paired modes can be classified into four types. The basic mechanisms for the four types of non-modal energy growths can be summarized, in terms of their initial modal cross-band kinetic energy X_j and along-band kinetic and buoyancy energy Y_j [see (6.2)], as follows:

(i) If $X_j > Y_j$ (or $X_j < Y_j$) for a pair of propagating modes, then the two modes can be combined to offset each other's cross-band velocity (or along-band velocity and buoyancy) and thus to minimize the total perturbation energy to $4Y_j$ (or $4X_j$) at the initial time. As the two modes propagate toward each other through one half of the wavelength (by one quarter of the wave oscillation time period), their associated cross-band velocity (or along-band velocity and buoyancy) fields become exactly in phase, so the total perturbation energy is increased to $4X_j$ (or $4Y_j$) and the non-modal growth reaches the maximum value of X_j/Y_j (or Y_j/X_j). The non-modal growth produced by paired propagating modes is classified as PP1 type if $X_j > Y_j$ or as PP2 type if $X_j < Y_j$. The PP1 growth is characterized by the increase of the cross-band kinetic energy that overly offsets the decrease of the along-band kinetic and buoyancy energy. The situation is opposite for the PP2 growth.

(ii) For a pair of growing and decaying modes, the two modes can be also combined to reduce the cross-band kinetic energy if $X_j > Y_j$ (or the along-band kinetic and buoyancy energy if $X_j < Y_j$) at the initial time and thus to enhance the growth of the total perturbation energy at the optimization time (see section 4b). In this case, the inclusion of the decaying mode reduces the total energy more at the initial time than at the optimization time, so the non-modal energy growth can be enhanced by a factor up to $(X_j + Y_j)^2/(4X_jY_j)$ as the optimization time approaches infinity. The non-modal growth produced by paired growing and decaying modes is classified as GD1 type if $X_j > Y_j$ or as GD2 type if $X_j < Y_j$.

When the optimization time is sufficiently short, the maximum non-modal growth is produced mainly by the paired fastest propagating modes (PP2 type). When the optimization time is large, the maximum non-modal growth is produced mainly by the paired slowest propagating modes (or fastest growing and decaying modes) if the parameter point (Ri, l) is outside (or inside) the unstable region. Outside the unstable region, the maximum non-modal growth is the PP1 type on the short-wavelength side with a local maximum (about 10) in the vicinity of $Ri = 1.1$, but changes to the PP2 type and becomes very large as $Ri < 1$ and l is immediately outside the unstable region (see Fig. 6c). Inside the unstable region, the maximum non-modal growth is the GD1 type on the short-wavelength side but changes to the GD2 type on the long-wavelength side immediately inside the unstable region.

The GD1 or GD2 non-modal growth is larger than the energy growth produced by the fastest growing mode, but the non-modal growth rate always approaches the constant modal growth rate as the optimization time increases. Unless the parameter point is immediately inside the unstable region (see Fig. 6c), the transient non-modal growth rate is not much larger than the modal growth rate and rapidly approaches the modal growth rate (within about two e-folding time periods). Because of this, the GD1 or GD2 non-modal growth (if it occurs) will play essentially the same role as the model growth in generating symmetric perturbations.

The PP1 non-modal energy growth produced by paired slowest propagating modes (with $l \leq 0.1$) is large (close to 10 for $\tau = 5$ as shown in Fig. 6c) when Ri is in the vicinity of 1.1. As this type of non-modal growth is characterized by the increase of the cross-band kinetic energy, it may generate strong cross-band vertical circulation over a wide range of optimization time. The PP2 non-modal energy growth produced by paired fastest propagating modes is not significant because the growth is small (between 1 and 2.4 for $\tau = 0.5$ as shown in Fig. 6a) and lasts only for a short time ($\tau < 1$). The PP2 non-modal energy growth produced by paired slowest propagating modes (with $l > 1$), however, can be very large and last for a long time, especially when the parameter point is near the unstable region. This type of non-modal growth is characterized by the increase of the along-band kinetic and buoyancy energy. Note that the vertical displacement (obtained by the time-integration of the vertical component of the cross-band velocity) is proportional to the along-band velocity and buoyancy, so this PP2 type of non-modal growth may provide a large vertical lift in the lower troposphere to trigger moist convection. Moist convection and convective storms can be triggered by propagating inertia-gravity waves in many different ways in the atmosphere and the related wave dynamics often appear to be approximately linear and more or less non-modal as suggested by observational studies (Uccellini 1975; Korch et al. 1988; Fovell et al. 2004). The energy norm used in this paper, however, does not directly measure the vertical displacement. To study the non-modal growth of the vertical displacement generated by inertia-gravity waves (including the propagating modes studied in this paper) in terms of triggering moist convection, a new metric needs to be introduced. This problem is under our investigation.

Acknowledgments. The author is grateful to Drs. Ting Lei and Shouting Gao for their numerical analyses (not shown in this paper) that stimulated and supported the analytical study in this paper, and to Dr. Robert Davies-Jones for proof-reading that improved the presentation. The work was supported by the NSF Grant ATM-9983077 to the University of Oklahoma.

References

Bennetts, D. A., and B. J. Hoskins, 1979: Conditional symmetric instability - possible explanation for frontal rainbands. *Quart. J. R. Met. Soc.*, **105**, 945-962.

Buizza, R. and Palmer, T. N., 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434-1456.

Emanuel, K. A., 1979: Inertial instability and mesoscale convective systems. Part I: Linear theory of inertial instability in rotating viscous fluids. *J. Atmos. Sci.*, **36**, 2425-2449.

Emanuel, K. A. 1983: The Lagrangian parcel dynamics of moist symmetric instability. *J. Atmos. Sci.*, **40**, 2368-2374.

Farrell, B. F., 1984: Modal and non-modal baroclinic waves. *J. Atmos. Sci.*, **41**, 668-673.

Farrell, B. F., and P. J. Ioannou, 1996: Generalized stability theory. I. Autonomous operators. *J. Atmos. Sci.*, **53**, 2025-2040.

Fovell, R., B. Rubin-Oster, and S.-H. Kim, 2004: A discretely propagating nocturnal Oklahoma squall line: Observations and numerical simulations. Preprints, *22th Conference on Severe Local Storms*, 4-8 Oct. 2004, Hyannis, Massachusetts, Amer. Meteor. Soc., Paper No. 6.1, Conference CD.

Koch, S. E., Golus R. E., and Dorian P. B., 1988: A mesoscale gravity wave event observed during CCOPE. Part II: Interactions between mesoscale convective systems and the antecedent waves. *Mon. Wea. Rev.*, **116**, 2545–2569.

Miller, T. L., 1985: On the energetics and nonhydrostatic aspects of symmetric baroclinic instability. *J. Atmos. Sci.*, **42**, 203-211.

Morgan, M. C., and C.-C. Chen, 2002: Diagnosis of Optimal Perturbation Evolution in the Eady Model. *Journal of the Atmospheric Sciences*: Vol. 59, No. 2, pp. 169–185.

Trefethen, L. N., A. E. Trefethen, S. C. Reddy, and T. A. Driscoll, 1993: Hydrodynamic stability without eigenvalues. *Science*, **216**, 578-584.

Uccellini, L. W., 1975: A case study of apparent gravity wave initiation of severe convective storms. *Mon. Wea. Rev.*, **103**, 497–513.

Xu, Q. 1986: Conditional symmetric instability and mesoscale rainbands. *Quart. J. R. Met. Soc.* , **112**, 315-334.

Xu, Q., 2004: Nearly symmetric and nearly baroclinic instabilities in the presence of diffusivity. Part II: Mode structures and energetics. *J. Fluid Mech.* **500**, 283-312.

Xu, Q., and X. P. Zhou, 1982: Baroclinic and inertial instability in an nonhydrostatic atmosphere. *Scientia Atmos. Sinica*, **6**, 355–367.