MET O 11 TECHNICAL NOTE NO. 162

SOME GENERAL PROPERTIES OF SOLUTIONS OF THE SEMIGEOSTROPHIC EQUATIONS INFERRED BY GEOMETRICAL CONSIDERATIONS

by

R. J. Purser* and M. J. P. Cullen

Meteorological Office (Met O 11)
London Road
Bracknell
Berkshire
United Kingdom

* On secondment at the Space Science and Engineering Laboratory, University of Wisconsin, Madison, Wisconsin 53706 (from 1st February 1983).

January 1983

JB

MET O 11 TECHNICAL NOTE NO 162

SOME GENERAL PROPERTIES OF SOLUTIONS OF THE SEMIGEOSTROPHIC EQUATIONS INFERRED BY GEOMETRICAL CONSIDERATIONS. By R J Purser and M J P Cullen.

Summary

The adiabatic semigeostrophic equations used by Hoskins and Bretherton (1972) and others provide a means whereby processes of frontogenesis may be idealized and studied using exact analytic solutions of these equations. Examination of the semigeostrophic equations reveals that there is a useful and natural correspondence between their dynamically stable solutions and the surfaces of a class of convex solids. Here we exploit this correspondence to derive a number of rather general properties of the two-dimensional semigeostrophic solutions deduceable from geometrical properties of convex solids. The results we obtain are not conditional on solutions being smooth. Some extensions of these results to three-dimensions are discussed.

1. INTRODUCTION

It is well known that in certain circumstances fluids behave in such a way as to generate in a finite time a sharp discontinuity in variables such as temperature and velocity from an initially smooth flow. The study of such behaviour in a rotating fluid is of obvious relevance to meteorologists when the result is the formation of a front, and the ability of conventional numerical models to simulate correctly the evolution of the resolvable part of the flow in such situations is a matter of special concern given the intrinsic importance of these features and the difficulty of handling their effects by finite difference methods. An approach to investigating the dynamical aspects of atmospheric and oceanic fronts that has been demonstrably successful in recent years is through the application of the semi-geostrophic equations. Many of the relevant results are presented in the paper of Hoskins and Bretherton (1972) (which we abbreviate to HB) where the development of discontinuities in a finite time is demonstrated explicitly for a number of configurations in which the special structure of the basic-state frontogenic deformation allows the solutions to be regarded as essentially two-dimensional.

Recently Cullen (1982) has extended the study of these equations, using explicitly Lagrangian techniques, to beyond the instant when a discontinuity forms and has shown how exact solutions can be constructed for a small system of simple finite elements in each of which the potential temperature and potential momentum are uniform. These solutions were then compared with the results obtained from the integration of a primitive equation model from equivalent initial data in order to reveal any significant shortcomings of the latter method as applied to frontal regions. From a mathematical point of view it is clearly desirable to be able to make some definite statements about the existence, multiplicity and other general properties of these reference-solutions to the two dimensional semigeostrophic set that are valid even after discontinuities have developed. This would then give credence to the results obtained by the algorithms that tackle those equations numerically in more complicated cases than those studied so far. This paper is concerned with such questions and will develop geometrical arguments to define some important general properties of possible solutions.

2. THE TWO-DIMENSIONAL FRONTOGENIC SEMI-GEOSTROPHIC EQUATIONS

In units in which $\theta_0 = f = g = 1$ the equations governing the frontogenesis models of HB become

$$\frac{DM}{Dt} + \alpha M = 0$$

$$\frac{D\theta}{Dt} = 0 \qquad (2.1)$$

$$\frac{DA}{Dt} + \alpha A = 0$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + w\frac{\partial}{\partial z} \quad ; \quad M \equiv v + x$$

M is called the 'potential momentum' and A is the area of any fluid element in the (x, z) plane. Also the x-component geostrophic equation and the hydrostatic equations apply:

$$\frac{\partial \phi}{\partial x} = v \quad ; \quad \frac{\partial \phi}{\partial z} = \theta \qquad (2.2)$$

We assume conditions of no flow across the prescribed boundary. Provided v and θ are differentiable a consequence of (2.1) is the conservation of potential vorticity:

$$\frac{Dq}{Dt} = 0$$

where $q = \dfrac{\partial(M, \theta)}{\partial(x, z)}$ (2.3)

It is useful to introduce a modified 'pressure' variable

$$\hat{\phi} = \phi + \frac{x^2}{2} \qquad (2.4)$$

So that

$$(M, \theta) = \left(\frac{\partial \hat{\phi}}{\partial x}, \frac{\partial \hat{\phi}}{\partial z}\right)$$

Then if

$$\underset{=}{Q} \equiv \begin{bmatrix} \dfrac{\partial^2 \hat{\phi}}{\partial x \partial x} & \dfrac{\partial^2 \hat{\phi}}{\partial x \partial z} \\ \dfrac{\partial^2 \hat{\phi}}{\partial z \partial x} & \dfrac{\partial^2 \hat{\phi}}{\partial z \partial z} \end{bmatrix} \qquad (2.5)$$

is defined, it follows that

$$q = \det(\underline{\underline{Q}}) \tag{2.6}$$

In the continuous equations the Hessian matrix Q is of special importance in determining the ageostrophic terms such as u, w and the pressure tendency, $\Gamma \equiv \frac{\partial \phi}{\partial t}$. For example, in order to determine Γ from the governing equations we would get

$$\left.\begin{aligned} &\underline{\nabla}\Gamma + \underline{\underline{Q}}\cdot\underline{u} = \underline{b} \\ \text{where}\quad &\underline{\nabla} \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial z}\right); \ \underline{u} \equiv (u, w); \ \underline{b} \equiv \left(-\alpha\frac{\partial\hat{\phi}}{\partial x}, 0\right) \end{aligned}\right\} \tag{2.7}$$

and use the continuity equation which, from (2.1), is:

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} + \alpha = 0 \tag{2.8}$$

to eliminate the velocity components:

$$\underline{\nabla}\cdot\left(\underline{\underline{Q}}^{-1}\cdot\underline{\nabla}\Gamma\right) = \alpha + \underline{\nabla}\cdot\underline{\underline{Q}}^{-1}\cdot\underline{b} \tag{2.9}$$

To solve equations of this type with appropriate boundary conditions (derived through substitution of known normal velocities into (2.7)) the differential operator on the left-hand side should be elliptic in character, implying that Q is positive-definite. In some circumstances (2.7) may be solved when Q is singular by using an integral form of the continuity equation (integrated along the line of the characteristics between opposite boundaries) but the problem is indisputably ill-posed when Q has a negative eigen-value. In the latter case the corresponding physical system would be subject to convective, inertial or symmetric instability and not properly describable in terms of the semi-geostrophic equations because acceleration components then become of comparable magnitude to the corresponding pressure gradient forces.

We may imagine $\hat{\phi}$ to be the 'elevation' of a 'surface' in a third coordinate normal to the (x, z) plane, in which case the matrix Q gives a measure of the 'curvature' components of the surface (in the limit in which the scale of elevation of $\hat{\phi}$ is negligible in comparison with the scale of x and z). Then q is essentially proportional to the Gaussian curvature of the surface. The condition that an eigenvalue of Q should nowhere be negative for a twice-differentiable distribution of $\hat{\phi}$ may be equivalently stated:

Definition

For a solution of the semigeostrophic equations, to be dynamically stable within the (x, z) - domain, D, any straight segment in $(x, z, \hat{\phi})$ space joining two points on the solution-surface, $\hat{\phi} = \tilde{\phi}(x,z)$, and whose projected (x, z) part lies wholly within D, must lie entirely above the solution-surface,

ie. $$\hat{\phi}(x, z) \geq \tilde{\phi}(x, z)$$

It is natural to expect this condition for stability to generalise to instances where $\tilde{\phi}$ is not everywhere differentiable. Returning again to the geometrical analogy, we are assuming that the solution-surface, $\tilde{\phi}(x, z)$, appears locally convex when viewed from underneath (where 'up' is the direction of increasing $\hat{\phi}$), i.e. the volume defined by $\hat{\phi} \geq \tilde{\phi}(x, z)$ is a convex subset of R^3. The geometrical notion of convexity is of sufficient importance here to justify a more formal discussion of it and its consequences.

Definition

A set S of points in the space R^n with a Euclidean metric is said to convex if, for all pairs, $\underset{\sim}{x}$, $\underset{\sim}{y} \in S$ the straight line segment between them is entirely in S.

The following important properties of convex sets are implied:

i. The intersection (in the set-theoretic sense) of convex sets is a convex set.

ii. A closed half space, $L(\underset{\sim}{x}) \geq 0$, is a convex set where L is a linear function of $\underset{\sim}{x}$.

iii. When a convex proper-subset, S, of R^n contains its boundary points, then S is the intersection of half-spaces of the type described in (ii).

Proof

Let x be any point not in S. There must be a unique 'nearest point', $\underset{\sim}{y} \in S$ (there cannot be two because their mid-point would be closer!). The half space of points Z satisfying.

$$(\underset{\sim}{z} - \underset{\sim}{y}) \cdot (\underset{\sim}{y} - \underset{\sim}{x}) \geqslant 0 \tag{2.10}$$

clearly contains S and excludes x. All points outside S may be excluded by the intersection of such half spaces.

Definition

The oriented hyperplane bounding the half-space of (2.10) is said to be a tangent-plane of S at y.

iv. To every point on the boundary of a convex set there is a tangent-plane that touches it. This property enables us to generalise the conventional definition of a 'gradient' to allow the (multiple) assignment of gradients to points of a convex surface even where differentiation is undefined.

v. Given orthogonal coordinates, x, ... x_n for points in R^n, let the 'prism', S_1 be the n-dimensional set,

$$(x_1, \ldots x_{n-1}) \in S_2$$

where S_2 is a compact closed subset of R^{n-1}. Let S_3 be a (non-empty) set of oriented half spaces (including their boundary points) of R^n for each of which the gradient,

$$\underset{\sim}{g} \equiv \left(\frac{\partial x_n}{\partial x_1}, \ldots \frac{\partial x_n}{\partial x_{n-1}} \right) \tag{2.11}$$

of the bounding hyperplane is finite and whose orientation is such as to include any point of fixed $(x_1 \ldots x_{n-1})$ coordinates if its x_n coordinate is sufficiently positive. By property (i) the intersection

$$S_4 = S_2 \cap S_3$$

is clearly convex. The set of tangent-planes to S_4 with finite gradients is 'complete' in the sense that every possible finite gradient as defined by (2.11) is represented. If S_5 is the 'lower' boundary of S_4 defined as the set $x = (x_1, \dots x_{n-1}, x_n) \in S_4$ such that for all $\delta > 0$

$$(x_1, \dots x_{n-1}, x_n - \delta) \notin S_4$$

Then there is a natural mapping from each point $\underset{\sim}{x} \in S_5$ to a (non-empty) set of (n-1)-dimensional gradients $\underset{\sim}{g}$ (i.e. those corresponding to the tangent planes of S_4 touching $\underset{\sim}{x}$). Conversely, there is a mapping from each gradient $\underset{\sim}{g}$ to a non-empty set of points in S_5 (i.e. those points which the tangent plane with gradient $\underset{\sim}{g}$ touches).

iv. The gradients at a point on S_5 form a set that is convex in gradient-space.

vii. For any continuous directed curve C in the set S_5 of (v) between end points $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ the mappings in (v) permit us to assign at least one continuous directed curve C' in the space of gradients such that each point $\underset{\sim}{x} \in C$ maps to a gradient $\underset{\sim}{g} \in C'$. The curves may be jointly parametrised by a continuous parameter $\sigma \in [a, b] \subset R^1$ such that

$$\underset{\sim}{x}(a) = \underset{\sim}{x}_1 \quad , \quad \underset{\sim}{x}(b) = \underset{\sim}{x}_2$$

and both $\underset{\sim}{x}$ and $\underset{\sim}{g}$ are continuous functions of σ with $\underset{\sim}{x} = \underset{\sim}{x}(\sigma)$ implying $\underset{\sim}{g}(\sigma)$ is the gradient of a tangent plane at $\underset{\sim}{x}$ (N.B. $\underset{\sim}{g}$ need not be a continuous or univalued function of x for this property to remain true).

If $\underset{\sim}{x}'$, the project of x, is a straight segment in the subspace $(x_1, \dots x_{n-1})$, for any pair $\underset{\sim}{x}(\sigma_2) \neq \underset{\sim}{x}(\sigma_1)$:

$$\left(\underset{\sim}{g}(\sigma_2) - \underset{\sim}{g}(\sigma_1)\right) \cdot \left(\underset{\sim}{x}'(\sigma_2) - \underset{\sim}{x}'(\sigma_1)\right) \geq 0 \tag{2.12}$$

In words, the forward component of gradient increases monotonically or is stationary along any straight line of $(x_1, \dots x_{n-1})$ in the set S_2 of (v).

viii. In the natural extension of result (2.12) to m-dimensions we state without proof the following generalization, using the definitions of the sets established in (vi):

An image, $G \subset \{\underset{\sim}{g}\}$ may be found under the gradient-mapping of any compact directed subset $A \subset S_2$ with $A \subset S_6 \equiv$ an m-dimensional subspace intersecting S_2, such that if the 'measure' of A is defined positive in S_6 then the measure of G also has non-negative magnitude in its orthogonal projection onto S_6.

By 'measure' we mean m-dimensional 'volume'. We make no attempt to justify this assertion rigorously. In the case of a two dimensional manifold S_5, i.e. a plane of S_2, the assertion for $m = 2$ is equivalent to the statement that a circuit on S_2 enclosing positive area corresponds to a closed curve in gradient space enclosing a non-negative area (there may also be an image-circuit in gradient-space enclosing negative area but in such a special case all the gradients enclosed by this part of the circuit are also images of a point on the circuit in S_2, so the direction may be reversed to recover a positive area).

The properties listed in (v), (vi), (vii) and (viii) are of particular interest when we make the correspondence

$$x_1 \equiv x, \; x_2 \equiv z, \; x_3 \equiv \hat{\phi} \quad \text{with} \quad m = 3$$

We shall deliberately limit ourselves to physical domains, D in (x, z)-space which are convex (but see remark 5.5). From the general definition of a dynamically stable solution and the definition of convexity it is clear that the set of points in R^3, $\hat{\phi}(x,z)$ satisfying,

$$\hat{\phi}(x,z) \geq \tilde{\phi}(x,z) \quad , \; (x,z) \in D \tag{2.13}$$

is convex.

3. EXISTENCE AND UNIQUENESS OF SOLUTIONS

Assume we are approximating solutions of (2.1) in the manner of Cullen (1982) using finite elements in which M and θ are uniform, i.e. the variation of $\hat{\phi}$ in each element is linear in x and z. The problem is specified by prescribing for each element i its area, A_i its potential momentum, M_i and its potential temperature θ_i together with the restriction that no two elements have both the same M and same θ. The solution-domain D is specified and assumed convex. Its area is A_D and must satisfy

$$A_D = \sum_{i=1}^{N} A_i \qquad (3.1)$$

where N is the number of elements. In the geometrical picture each element is associated with a flat facet of a polyhedral shell spanning domain D. The location of this shell is uniquely determined given the gradient components M, θ and the coordinate of the associated tangent plane at some reference location, say $(x, z) = (0, 0)$. To show this, let $\tilde{\Phi}_i(x, z, \phi'_i)$ be the $\hat{\Phi}$ of the plane containing element i with gradient (M_i, θ_i) and satisfying $\tilde{\Phi}_i(0, 0, \phi'_i) = \phi'_i$ The shell is then the surface $\tilde{\Phi}(x, z, \underset{\sim}{\phi'})$ such that for each fixed $x, z, \underset{\sim}{\phi'}$:

$$\left.\begin{aligned} &\tilde{\Phi}(x, z, \underset{\sim}{\phi'}) = \max_i \left\{ \tilde{\Phi}_i(x, z, \phi'_i) \right\} \\ \text{where} \quad &\underset{\sim}{\phi'} \equiv (\phi'_1, \phi'_2 \cdots) \end{aligned}\right\} \qquad (3.2).$$

This is illustrated schematically in the 'cross-section' of Figure 1. By property (i) of Section 2 each facet of any shell constructed in this way is itself a convex region and is therefore simply connected, and each edge between facets is a striaght segment. Consider an arbitrary arrangement of the set of element-planes specified by the coordinates, $\{M_i, \theta_i, \phi'_i\}$ in which a particular element, say K, has finite area, A_K

$$0 < A_K < A_D$$

then,

<u>Lemma</u> 3.1

Any increase in ϕ'_K with the other coordinates, $\phi'_i, i \neq K$ and M_i, θ_i kept constant will result in an increase in the area A_K balanced by a decrease in the areas of each of the neighbouring facets.

Proof

The normal component of the change in gradient across the boundary between neighbouring elements K and L is

$$\left(\frac{\partial \hat{\phi}_L}{\partial n} - \frac{\partial \hat{\phi}_K}{\partial n} \right) = \Delta \frac{\partial \hat{\phi}_{KL}}{\partial n} > 0$$

where n is the distance in the direction pointing outwards from K. It is strictly a positive change in gradient because the set of points defined by (2.13) would not be convex otherwise. Increasing ϕ'_K by amount $\delta\phi'_K$ moves this boundary by a distance

$$d = \frac{\delta \phi'_K}{\Delta \frac{\partial \hat{\phi}_{KL}}{\partial n}}$$

and therefore increases the area of element K at the expense of the area of L and K's other neighbours. Figure 2 shows this.

Theorem 3.2

Given the intended slopes and areas $\{M_i, \Theta_i, \tilde{A}_i\}$ of the set of elements, then if a solution characterised by the set $\{\phi'_i\}$ exists it is unique within a uniform change in all the ϕ'.

Proof

Let $T_1 = \{\phi_i'^{(1)}\}$ be one solution. Consider the hypothesis that another solution $T_2 = \{\phi_i'^{(2)}\}$ also satisfies the equations. Divide the set of elements into those E_a, for which the difference

$$\phi_i'^{(2)} - \phi_i'^{(1)} = F$$

is the maximum, and the remainder, E_b. We transform solution T_1 to solution T_2 in two stages: first add F to the ϕ' of both E_a and E_b to form T_1'. Clearly this does not materially affect the solution (areas of elements remain the same). The step from T_1' to T_2 now requires only negative changes in ϕ' to elements of the set E_b. However, we showed that the area of an element changes monotonically with its ϕ' at the expense of its neighbours. Thus if E_b is non-empty at least one element of it must be in contact with an element in E_a and hence the area of an element in E_a will be increased, contrary to the hypothesis. Hence E_b is empty and the theorem is proved.

Theorem 3.3

Given slopes and areas $\{M_i, \theta_i, \tilde{A}_i\}$ with $\sum \tilde{A}_i = A_D$ a solution $\{\phi'_i\}$ exists.

Proof

A 'feasible' solution is any surface of the type (3.2) formed from planes possessing the correct slopes $\{M_i, \theta_i\}$ but not necessarily yielding elements with the correct areas. Each feasible solution is determined by the set $\{\phi'_i\}$ and we may assume an arbitrary solution of the form (3.2) as a first guess. In the proof that follows we shall assume that the gradients $\{M_i, \theta_i\}$ remain correct and we iterate towards a set $\{\phi'_i\}$ with the correct areas. Then for a given set of ϕ' we define the error norm:

$$N_2(\underset{\sim}{\phi'}) = \sum_i |A_i - \tilde{A}_i|^2$$

(Each A_i is a function of all the ϕ'_j). Within a uniform change in ϕ', all possible combinations of intersections of the elementary planes can be obtained in a finite range of ϕ' so within this range N_2 attains its minimum.

Let
$$\varepsilon_i = A_i - \tilde{A}_i \tag{3.3}$$

Divide the elements into two sets: E_a contains elements for which ε is the maximum, ε_a; E_b contains the remainder. Suppose N_2 attains its minimum. Then if E_2 is not empty at least one member, k, of E_a borders members of E_b. Element k may also border other members of E_a. By reducing ϕ'_K we may reduce A_K by an arbitrary amount, say (a+b). By doing so we assume we increase the combined areas of other elements of E_a by a and of elements in E_b by b. The resulting change in the norm N_2 comprises three parts.

i. $\Delta N_2^{(a)}$ due to changes in areas of elements in E_a excluding K.

ii. $\Delta N_2^{(k)}$ the change due to area k.

iii. $\Delta N_2^{(b)}$ the contribution from set E_b.

Let $\mathcal{E}_b$ be the maximum error for members of E_b ($\mathcal{E}_b < \mathcal{E}_a$). Then using the simple inequality of the form:

$$\left(\sum (A_i + a_i)^2 - \sum A_i^2\right) \leqslant \left(\max_i A_i + \sum a_i\right)^2 - \left(\max_i A_i\right)^2 \tag{3.4}$$

such that $a_i > 0$

It follows that

$$\Delta N_2^{(a)} \leqslant 2\,\mathcal{E}_a\, a + a^2$$

$$\Delta N_2^{(b)} \leqslant 2\,\mathcal{E}_b\, b + b^2$$

$$\Delta N_2^{(k)} = -2\,\mathcal{E}_a (a+b) + a^2 + b^2 + 2ab$$

hence

$$\Delta N_2 \leqslant -2(\mathcal{E}_a - \mathcal{E}_b)b + 2(a^2 + b^2 + ab) \tag{3.5}$$

Since the rate at which an element's area changes with respect to its ϕ' is bounded, there is always a sufficiently small change in ϕ', and hence in b and a such that

$$(\mathcal{E}_a - \mathcal{E}_b)b > (a^2 + b^2 + ab) \tag{3.6}$$

i.e. $\Delta N_2 < 0$

Therefore E_b must be a null-set and the errors of areas are equals. But since they sum to zero these errors vanish when N_2 attains its minimum. Thus a solution exists.

The conditions of existence and uniqueness established for the discrete case are not by themselves sufficient to justify the same assertions for solutions in which the distribution of physical-area per unit area of gradient-space (which is the inverse of potential vorticity) is generalised to allow continuous variation in the gradient-space, nor to cases in which the number of facets is countably infinite. This is required to complete the existence proof for the equations (2.1). For this it is desirable to demonstrate that solutions possess a certain degree of continuity with respect to any variations in the gradients. At present the required conditions have not been proved but we shall nevertheless outline a plausible way by which the necessary connection between the finite discrete cases and the general cases might be made.

The discrete problem may be expressed in terms of an impulsive distribution of the quantity, $(1/q)$, in gradient space composed of a finite number of delta-functions. Each delta-function is associated with a facet of the solution whose area is the coefficient of the delta function and whose gradient is given by its location in gradient-space. The integral of the distribution gives the physical area of the solution domain, D. Within a compact domain of gradient space we may associate with any non-negative distribution $G^{\infty}(\underset{\sim}{g})$ of $(1/q)$ (including impulsive components) a sequence of finite discrete approximations, $G^{(k)}(\underset{\sim}{g})$, to $G^{(\infty)}(\underset{\sim}{g})$ with $k \geq 1$ in the following unambiguous way:

Compose $G^{(k)}(\underset{\sim}{g})$ with a regular array of weighted delta-functions:

$$G^{(k)}(M, \theta) = \sum_{ij} C_{ij}^{(k)} \, \delta\left(M - (i)\Delta g(2^{-k}), \theta - (j)\Delta g(2^{-k})\right) \tag{3.7}$$

where $C_{ij}^{(k)}$ is the total physical area of the distribution $G^{(\infty)}(\underset{\sim}{g})$ within the set,

$$S_{ij}^{(k)} \equiv \left\{M, \theta \mid (i)\Delta g(2^{-k}) \leq M < (i+1)\Delta g(2^{-k}), (j)\Delta g(2^{-k}) \leq \theta < (j+1)\Delta g(2^{-k})\right\} \tag{3.8}$$

the delta-function, $\delta(\underset{\sim}{g})$, of (3.7) satisfies the usual definition:

$$\delta(\underset{\sim}{g}) = 0 \qquad \underset{\sim}{g} \neq 0 \qquad \text{but} \qquad \iint \delta(\underset{\sim}{g}) = 1$$

The subsets $S_{ij}^{(k)}$ for a given k are disjoint and completely cover the gradient-domain. For $k > 1$ the distribution $G^{(k)}$ may be transformed to $G^{(k-1)}$

by translating each impulsive element in gradient-space keeping the intensity constant. This translation (which corresponds to a change in the shape of each facet keeping their areas constant) is by an amount less than or, equal to,

$$\delta g^{(k)} = \frac{\Delta g}{\sqrt{2}}(2^{-k}) \tag{3.9}$$

as may be seen in Figure 3. If it can be shown that relative differences in the solution surface between any fixed pair of points in the physical domain D during this transformation are always bounded by

$$|\delta\tilde{\phi}| < \alpha\,\delta g^{(k)} \tag{3.10}$$

where α is not a function of k, then it is clear that the solutions $\tilde{\phi}^{(k)}$ corresponding to $G^{(k)}$ form a sequence that converges uniformly to a limit that we may consistently define to be the solution $\tilde{\phi}^{(\infty)}$ corresponding to the original distribution $G^{(\infty)}$. A proof of (3.10) would be sufficient, though not perhaps necessary, to extend the theorems on existence and uniqueness to any compact non-negative distribution $G^{(\infty)}$. (3.10) is plausible because the perturbations in slope cause a small change in $\tilde{\phi}$ unless the facets have to be rearranged to ensure convexity. If they do have to be rearranged, the facets involved must all have similar slope and the effect on $\tilde{\phi}$ will still be small. It is trivially true in one-dimension, as illustrated in Figure 4 with α equal to the length of the domain.

4. THREE-DIMENSIONAL SEMI-GEOSTROPHIC THEORY

The three-dimensional equations are the natural generalization of the two dimensional set we have studied so far. The potential momentum M generalises to the 'geostrophic coordinates', X and Y, which change non-trivially in time. A detailed study of the three-dimensional equations is given in Hoskins and Draghici (1977). For our purposes the essential equations are:

$$\left.\begin{aligned} \frac{Du_g}{Dt} - v_{ag} &= 0 \\ \frac{Dv_g}{Dt} + u_{ag} &= 0 \\ \frac{D\theta}{Dt} &= 0 \end{aligned}\right\} \tag{4.1}$$

with

$$\left.\begin{aligned} &\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \\ &v_g = \frac{\partial \phi}{\partial x} \; ; \; -u_g = \frac{\partial \phi}{\partial y} \; ; \; \theta = \frac{\partial \phi}{\partial z} \end{aligned}\right\} \tag{4.2}$$

and

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + (u_g + u_{ag})\frac{\partial}{\partial x} + (v_g + v_{ag})\frac{\partial}{\partial y} + w\frac{\partial}{\partial z} \tag{4.3}$$

with

$$u = u_g + u_{ag} \quad , \quad v = v_g + v_{ag} .$$

Defining,

$$X = x + v_g \quad , \quad Y = y - u_g \tag{4.4}$$

then

$$\frac{DX}{Dt} = u_g \quad ; \quad \frac{DY}{Dt} = v_g \tag{4.5}$$

Rewriting as

$$\left.\begin{aligned} \frac{\partial X}{\partial t} + \underset{\sim}{u_g} \cdot \underset{\sim}{\nabla} X + \underset{\sim}{u_{ag}} \cdot \underset{\sim}{\nabla} X &= u_g \\ \frac{\partial Y}{\partial t} + \underset{\sim}{u_g} \cdot \underset{\sim}{\nabla} Y + \underset{\sim}{u_{ag}} \cdot \underset{\sim}{\nabla} Y &= v_g . \\ \frac{\partial \theta}{\partial t} + \underset{\sim}{u_g} \cdot \nabla \theta + \underset{\sim}{u_{ag}} \cdot \underset{\sim}{\nabla} \theta &= 0 \end{aligned}\right\} \tag{4.6}$$

where

$$\underset{\sim}{u_g} \equiv (u_g, v_g, 0) \; ; \; \underset{\sim}{u_{ag}} \equiv (u_{ag}, v_{ag}, w) \tag{4.7}$$

It is natural to solve this set by a splitting method. The terms $\underset{\sim}{U_g} \cdot \underset{\sim}{\nabla}$ are straightforward to evaluate and unlikely to cause existence problems provided $\underset{\sim}{U_g}$ remains bounded. We concentrate on $\underset{\sim}{U_{ag}} \cdot \underset{\sim}{\nabla}$ terms which are really implicit equations for $\underset{\sim}{U_{ag}}$ since (X, Y, θ) must satisfy the definition,

$$\left.\begin{aligned} (X - x) &= \frac{\partial \phi}{\partial x} \\ (Y - y) &= \frac{\partial \phi}{\partial y} \\ \theta &= \frac{\partial \phi}{\partial z} \end{aligned}\right\} \quad (4.8)$$

To isolate this problem we consider the following sub-problem of (4.6)

$$\left.\begin{aligned} \underset{\sim}{U_{ag}} \cdot \underset{\sim}{\nabla} X &= A \\ \underset{\sim}{U_{ag}} \cdot \underset{\sim}{\nabla} Y &= B \\ \underset{\sim}{U_{ag}} \cdot \underset{\sim}{\nabla} \theta &= C \\ \underset{\sim}{\nabla} \cdot \underset{\sim}{U_{ag}} &= 0 \end{aligned}\right\} \quad (4.9)$$

A, B, C given

then we seek a solution of (4.9) for $\underset{\sim}{U_{ag}}$ such that (X, Y, θ) satisfy (4.8). If existence of this solution can be proved it is likely that it can be extended to a proof of the existence of a solution of (4.6) by a splitting technique.

The natural definition of potential vorticity, generalizing equation (2.3), is

$$q = \frac{\partial(Y, -X, \theta)}{\partial(x, y, z)} \quad (4.10)$$

with the conservation law,

$$\frac{Dq}{Dt} = 0 \quad (4.11)$$

Let

$$\hat{\phi} = \phi + \tfrac{1}{2}(x^2 + y^2) \quad (4.12)$$

then

$$X = \frac{\partial \hat{\phi}}{\partial x} \; ; \; Y = \frac{\partial \hat{\phi}}{\partial y} \; , \; \theta = \frac{\partial \hat{\phi}}{\partial z} \quad (4.13)$$

Hence

$$q = \det(\underset{\approx}{Q})$$

where $$\underset{\approx}{Q} \equiv \begin{bmatrix} \frac{\partial^2\hat{\phi}}{\partial x\partial x} & \frac{\partial^2\hat{\phi}}{\partial x\partial y} & \frac{\partial^2\hat{\phi}}{\partial x\partial z} \\ \frac{\partial^2\hat{\phi}}{\partial y\partial x} & \frac{\partial^2\hat{\phi}}{\partial y\partial y} & \frac{\partial^2\hat{\phi}}{\partial y\partial z} \\ \frac{\partial^2\hat{\phi}}{\partial z\partial x} & \frac{\partial^2\hat{\phi}}{\partial z\partial y} & \frac{\partial^2\hat{\phi}}{\partial z\partial z} \end{bmatrix} \qquad (4.14)$$

Just as in the two dimensional equations the elements of the inverse of matrix Q determine the coefficients of the equation for the principal ageostrophic quantities. Therefore, as in the two-dimensional case, dynamical stability of the system for twice differentiable Q requires that Q nowhere has a negative eigenvalue. For a general (not necessarily smooth) solution, $\hat{\phi} = \tilde{\phi}(x, y, z)$, the condition for stability may be formally stated:

Definition

For a solution of the three-dimensional semi-geostrophic equations to be dynamically stable within the (x, y, z) - domain, D, any straight segment in $(x, y, z, \hat{\phi})$ - space joining two points on the solution manifold $\hat{\phi} = \tilde{\phi}(x, y, z)$ and whose projected (x, y, z) part lies wholly within D must be entirely within the four-dimensional region, $\hat{\phi}(x,y,z) \geq \tilde{\phi}(x,y,z)$.

The arguments concerning the convexity of the region of $(\underset{\sim}{x}, \hat{\phi})$ bounded (toward negative $\hat{\phi}$) by the solution manifold, $\tilde{\phi}$, that we applied in section 3 will follow also in three dimensions so that:

Theorem 4.1

Inside a convex three-dimensional domain, D, there is a unique arrangement of a given finite collection of elements of the fluid, each with a uniform prescribed 'gradient', (X_i, Y_i, θ_i) and volume, τ_i, that is dynamically stable provided that

$$\sum_i \tau_i = \tau_D$$

where τ_D is the volume of domain D.

The proof is identical to the proofs of theorems (3.2) and (3.3). This theorem as yet remains unproved for the continuous case (although it seems plausible) in which an arbitrary non-negative (but possibly impulsive) distribution of $(1/q)$ is prescribed in (X, Y, Θ) - space. A rigorous proof of this result would immediately verify the existence of solutions for the adjustment problem (4.9).

5. FURTHER RESULTS

The motivation for studying the semi-geostrophic equations has been primarily to simplify the dynamics of frontogenesis to their bare essentials. This section will discuss some general properties of the semi-geostrophic equations, resulting from the arguments of convexity that are related to the formation or existence of discontinuities in two-dimensions.

Theorem 5.1

If the potential vorticity is bounded, solutions to the semi-geostrophic equations cannot possess localized internal discontinuities in M or θ . Any discontinuities necessarily extend along an entire chord of the domain.

Proof

Suppose at some point the gradient of the solution $\tilde{\phi}$ changes discontinuously. Because of the symmetry in the equations we may for convenience consider a coordinate translation and rotation and a redefining of $\tilde{\phi}$ by the addition of a linear function of position so that the discontinuity occurs at $\underset{\sim}{x}_0 \doteq (0,0)$ with gradients

$$\underset{\sim}{g}_- = (-g, 0) \quad ; \quad \underset{\sim}{g}_+ = (g, 0) \quad \text{and} \quad \tilde{\phi}(\underset{\sim}{x}_0) = 0$$

By the convexity property (vi) of section (2) the gradients of point $\underset{\sim}{x}_0$ occupy only the straight segment ($\underset{\sim}{g}_-$, $\underset{\sim}{g}_+$). (The only convex sets of zero area are line segments).

By convexity the solution $\tilde{\phi}$ satisfies

$$\left.\begin{aligned} &\tilde{\phi}(x, z) \geqslant -xg \qquad x \leq 0 \\ &\tilde{\phi}(x, z) \geqslant xg \qquad x \geqslant 0 \end{aligned}\right\} \tag{5.1}$$

and the gradient $\underset{\sim}{g}$ at $\underset{\sim}{x}$ must satisfy

$$\underset{\sim}{g} \cdot \underset{\sim}{x} \geqslant \tilde{\phi}(\underset{\sim}{x}) \tag{5.2}$$

Since otherwise $\tilde{\Phi}(\underset{\sim}{x_0})$ would be less than the $\hat{\Phi}$ at x_0 of the tangent-plane touching x. Consider the rectangular circuit formed by the vertices:

A $\equiv (-\delta, 0)$

B $\equiv (-\delta, l)$

C $\equiv (\delta, l)$

D $\equiv (\delta, 0)$

as shown in Figure 5.

Suppose that on segment BC

$$\tilde{\Phi}(x) \geqslant g\delta \tag{5.3}$$

Then at each point on the circuit, inequality (5.2) defines a half space from which the possible gradients are excluded in gradient-space. The intersection of these excluded regions for the entire circuit is the triangular area in gradient space whose corners are:

$\alpha \equiv (-g, 0)$

$\beta \equiv (0, g\frac{\delta}{l})$

$\gamma \equiv (g, 0)$

as illustrated in Figure 6. The potential circulation inside this triangular region is:

$$C(\delta, l) = g^2 \frac{\delta}{l} \tag{5.4}$$

Since the potential vorticity everywhere is bounded, say by q_{max}, and the area of circuit ABCD is $l\delta$, then

$$q \leq q_{max} \Rightarrow l^2 \geqslant \frac{g^2}{q_{max}} \quad \text{for all} \quad \delta \tag{5.5}$$

Conversely, for any l, with $|l| < g/\sqrt{q_{max}}$

$$q \leq q_{max} \Rightarrow \tilde{\Phi}(\underset{\sim}{x}) \leqslant g\delta \quad \text{somewhere on BC for all} \quad \delta \tag{5.6}$$

Hence

$$\tilde{\Phi}(0, l) = 0 \qquad |l| < g/\sqrt{q_{max}} \tag{5.7}$$

(5.1) and (5.7) together imply that the discontinuity in gradient extends along this line segment between $(0, -g/\sqrt{q_{max}})$ and $(0, g/\sqrt{q_{max}})$. The argument can then be applied near the ends of this line segment to continue it as far as the boundary, thus completing the proof.

Analogous constructions may be made in the three-dimensional equations of Hoskins and Draghici (1977), initially by replacing circuit ABCD by the surface of a disc formed by rotating ABCD about AD. The corresponding excluded domain of gradient-space is the double-cone formed by rotating triangle $\alpha\beta\gamma$ about $\alpha\gamma$.

This theorem is a stronger statement than the statement in HB (p. 16) as it prohibits fronts even at a finite discontinuity of potential vorticity.

Definition

The 'convex-hull' of a set S is the minimal convex set Sc containing all the points of S. The 'convex-hull' of a distribution is the minimal convex set Sc such that the integral of the distribution over any region outside S_c vanishes. This concept is illustrated in Figure 7.

Theorem 5.2

Elements of the fluid with gradients on the convex-hull of the distribution of $(1/q)$ in gradient-space remain in contact with the boundary of the convex physical domain, D.

Proof

For any element, e, with gradient, $\underset{\sim}{g}_e$, at the boundary of the convex-hull of the distribution of $(1/q)$, then any outward vector, $\underset{\sim}{n}$, normal to a tangent there is such that this element has the maximum possible component of gradient in this direction. Take any point $\underset{\sim}{x}_e \in e$ and consider the line,

$$\underset{\sim}{x} = \underset{\sim}{x}_e + d\underset{\sim}{n}$$

on the tangent-plane to the solution-surface touching $\underset{\sim}{x}_e$. The solution-surface cannot depart from this line for positive a since this would require a gradient component in the direction of $\underset{\sim}{n}$ exceeding the maximum. Thus element e is in contact with the boundary.

The process of frontogenesis involves the pinching together of boundary points of the fluid usually followed by the intrusion of the resulting contact discontinuity into the interior of the domain. When the initial potential vorticity is bounded the possible locations of onset of front formation are restricted by the following theorem:

Theorem 5.3

Given that the initial distribution of gradient, (M,θ) is continuous with respect to $\underset{\sim}{x}$ and that the potential vorticity, q , is bounded, then if the region, of gradient-space where $(1/q)$ is strictly positive is delineated by the closed curve, C_G , a front can only form at a position corresponding to a concave portion of C_G.

Proof

Under the gradient mapping, the image of the open set of points strictly inside the domain D is initially the set of gradients strictly inside G_D. By a corollary of theorem 4.1, when a discontinuity has formed the potential vorticity in its neighbourhood is no longer bounded so it may be concluded that the image of points strictly inside D now contains gradients that are outside G_D. The extra gradients clearly belong to points of the discontinuity since these are the only points that are new to the interior of D. The extra gradients belonging to a particular point, $\underset{\sim}{x}_f$, of the front lie on the straight segment joining a pair of gradients g_1 and g_2 of the curve C_G, where g_1 , g_2 are gradients on each side of the front at $\underset{\sim}{x}_f$. The onset of development of an intruding discontinuity is therefore associated with a concave portion of C_G.

This last theorem generalises the observation of Hoskins (1971) (p 143) that in studies using the $\tan^{-1}(X)$ profile of potential temperature the surface front forms in the warmer half of the fluid and the upper front forms in the colder half of the fluid as shown schematically in Figure 8.

Remark 5.4

An important and rather paradoxical feature of these frontogensis models concerns the total potential circulation, C, associated with the entire cross-section of the fluid. From a glance at equations (2.1) one would expect C to obey the modified conservation law,

$$\frac{D}{Dt}\left(C(t)\, e^{\alpha t} \right) = 0 \tag{5.8}$$

and this would certainly be the case if the circuit in (x, z)-space associated with C were purely advected. However, if C is associated with a circuit at the domain boundary, equation (5.8) is no longer obeyed after a front forms because its development is associated with the injection of a new region of (M, θ)-space (i.e. the shaded area of Figure 8b) into the fluid interior as an impulsive line-source of potential vorticity. Thus while the potential vorticity is conserved followin each fluid-parcel the mean potential-vorticity may actually change!

Remark 5.5

The uniqueness theorem 3.2 was shown to hold for finite-element solutions at the semi-geostrophic equations in a convex domain. It is worth remarking that uniqueness cannot generally be guaranteed in a non-convex domain because concave regions of the boundary can separate distinct elements of the fluid with the same M and θ. A continuum of possible solutions exist according to the apportioning of the total area of fluid with given M and θ between the two or more distinct regions. The principle is illustrated in Figure 9.

Remark 5.6

The solutions we have considered are those corresponding to dynamically stable conditions and, under certain restrictions, these solutions have been shown to be unique. They can formally be found from an initially unstable configuration of the fluid by a rearrangement of the fluid elements. However, in these 'unbalanced' conditions the assumptions that restrict the validity of the semi-geostrophic approximation no longer apply so the consequent rearrangement of fluid elements is unlikely to proceed realistically.

6. CONCLUSION

In this paper we have shown that the definition of dynamical stability in the semi-geostrophic equations in two or three dimensions may be interpreted as a statement of the convexity of an associated region in three or four dimensions. This correspondence allows some general properties of semigeostrophic solutions to be inferred from results in geometry. It has not yet been possible to prove the existence and uniqueness of continuous solutions in a convex domain but a tentative approach has been suggested by which the desired proof might be obtained.

References

Cullen, M. J. P. 1982 Solutions to a model of a front forced by deformation. (Unpublished) Met O 11 Technical Note No. 160.

Hoskins, B. J. 1971 Atmospheric frontogenesis models: some solutions, Quart. J. R. Met. Soc., 97, pp 139-153.

Hoskins, B. J. and Bretherton, F. P. 1972 Atmospheric frontogenesis models: Mathematical formulation and solution. J. Atmos. Sci. 29, pp 11-37.

Hoskins, B. J. and Draghici, I. 1977 The forcing of ageostrophic motion according to the semi-geostrophic equations and in an isentropic coordinate model. J. Atmos. Sci. 34, pp 1859-1867.

Figure Captions

Figure 1. Schematic cross-section of solution-surface, showing its composition from the elementary planes.

Figure 2. (a) Plan and (b) Cross-section illustrating Lemma 3.1.

Figure 3. Plan view showing locations in gradient (M, θ)-space of impulsive components of distributions $G^{(K)}$ and $G^{(K-1)}$, and the region of this space (shaded) associated with the particular impulsive element of $G^{(K-1)}$ with coefficient $C_{ij}^{(K-1)}$.

Figure 4. Schematic illustration of (3.10) in one-dimension with a bounded change in the gradients of 'facets' (without change in their lengths) and α given by the span of the domain.

Figure 5. The circuit in (x, z)-space for theorem (5.1).

Figure 6. The triangle $\alpha\beta\gamma$ of gradient-space from which gradients on the circuit ABCD of Figure 5 referred to by theorem (5.1) are excluded.

Figure 7. (a) An arbitrary set, S, and (b) its convex-hull, Sc.

Figure 8. Schematic illustration of a stage in the growth of the frontal discontinuities for a uniform potential-vorticity, $\tan^{-1}(X)$ temperature-profile case, showing the correspondence between points a, b, c, d, e in (x, z)-space (a) and a', b', c', d', e' in (M, θ)-space (b). The shaded regions of (b) correspond to the frontal discontinuities.

Figure 9. A particular example of the non-uniqueness of solutions when the solution - domain is not convex. M is uniform, θ takes either of two values: θ_1 (shaded) or θ_2 ($>\theta_1$).

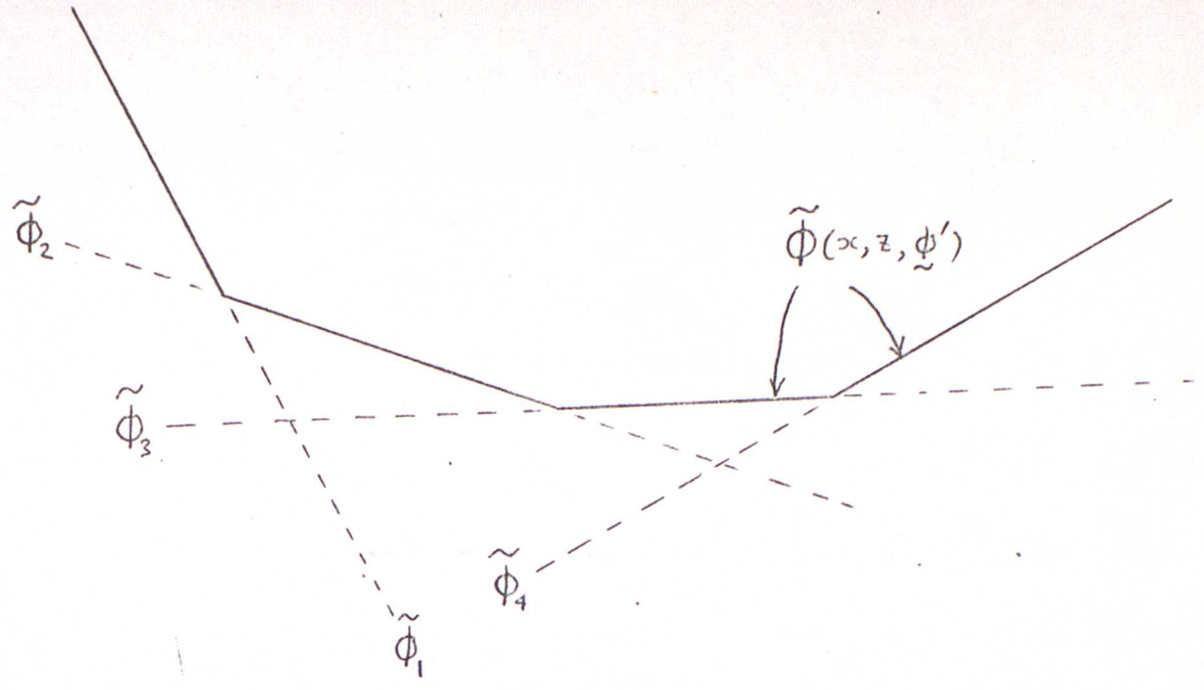


Figure 1

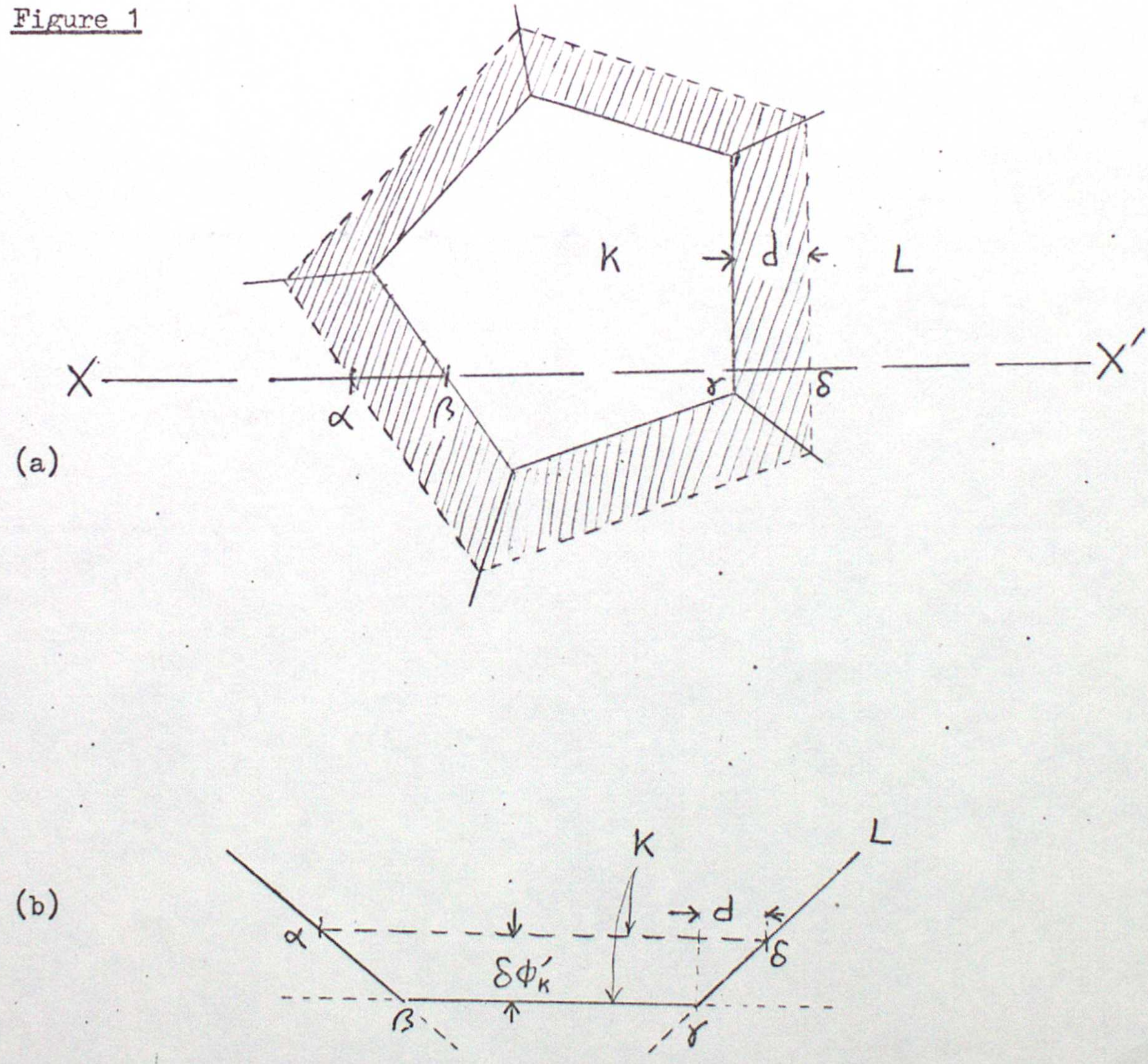
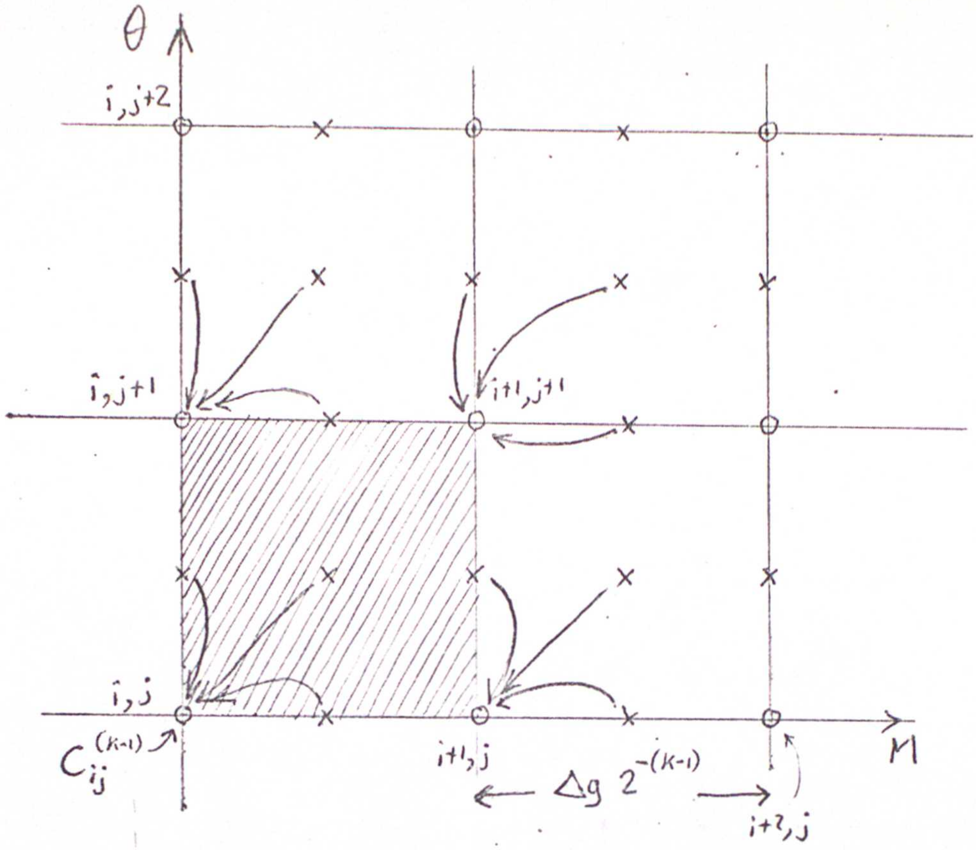


Figure 2



x Locations of points of distribution, $G^{(k)}$

o Locations of points also in $G^{(k-1)}$

▨ Region of gradient space associated with $C_{ij}^{(k-1)}$

<u>Figure 3</u>

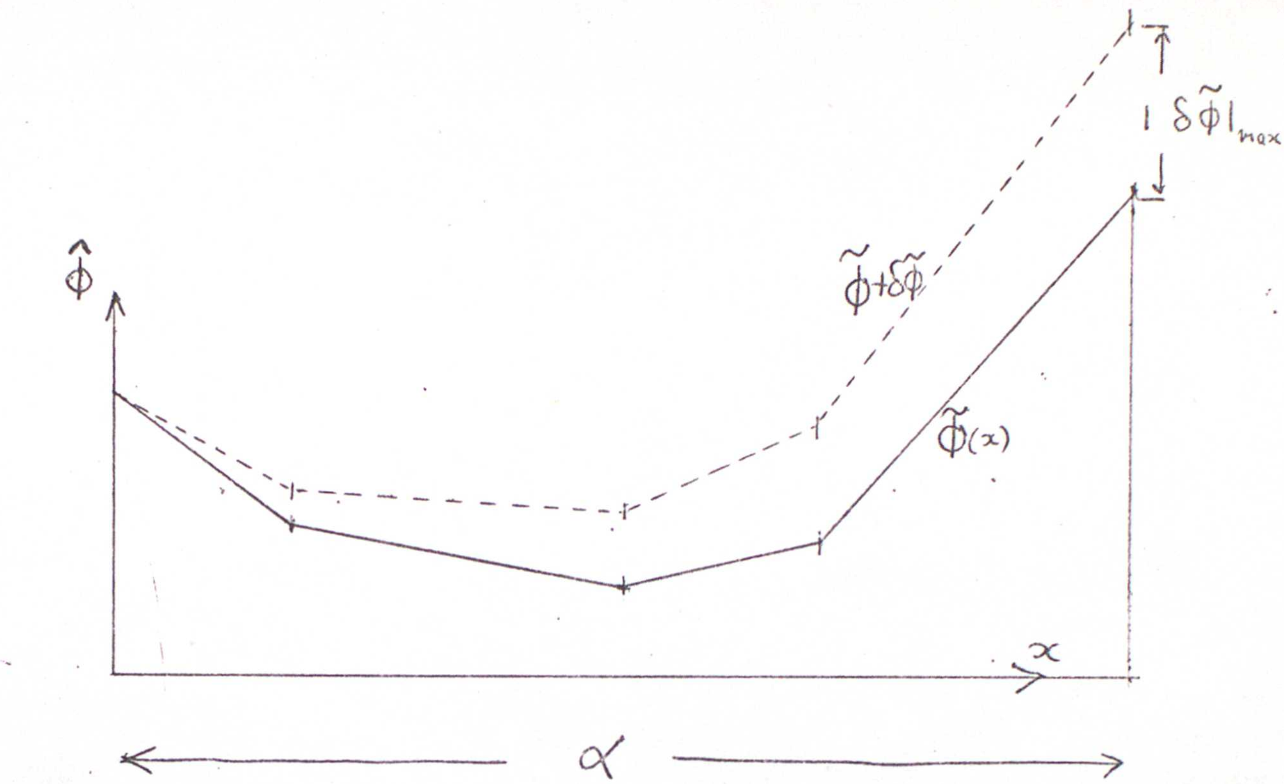


Figure 4

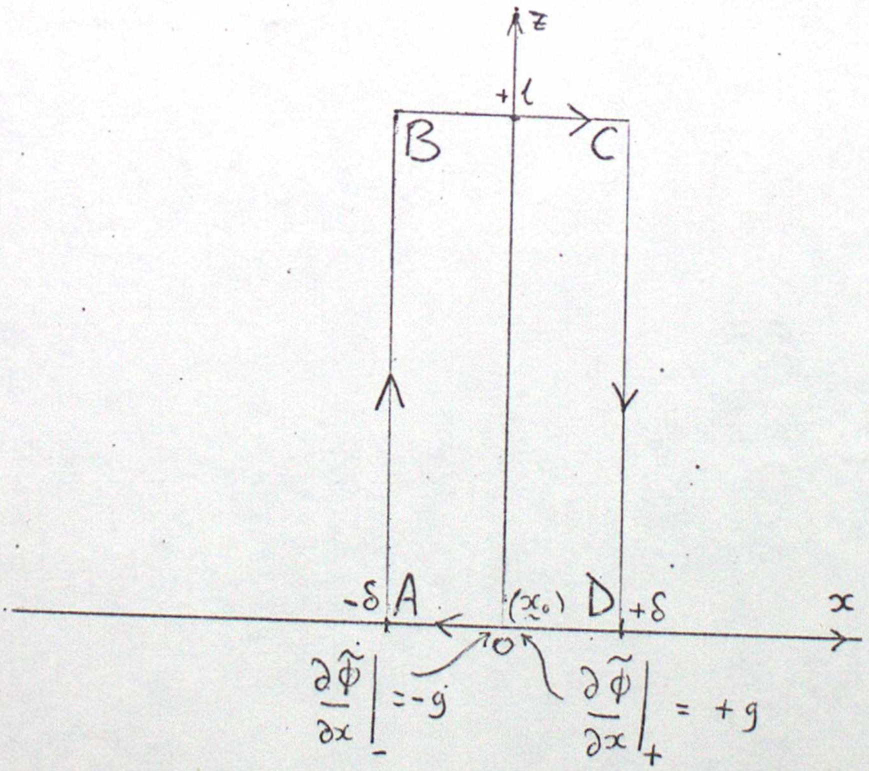


Figure 5

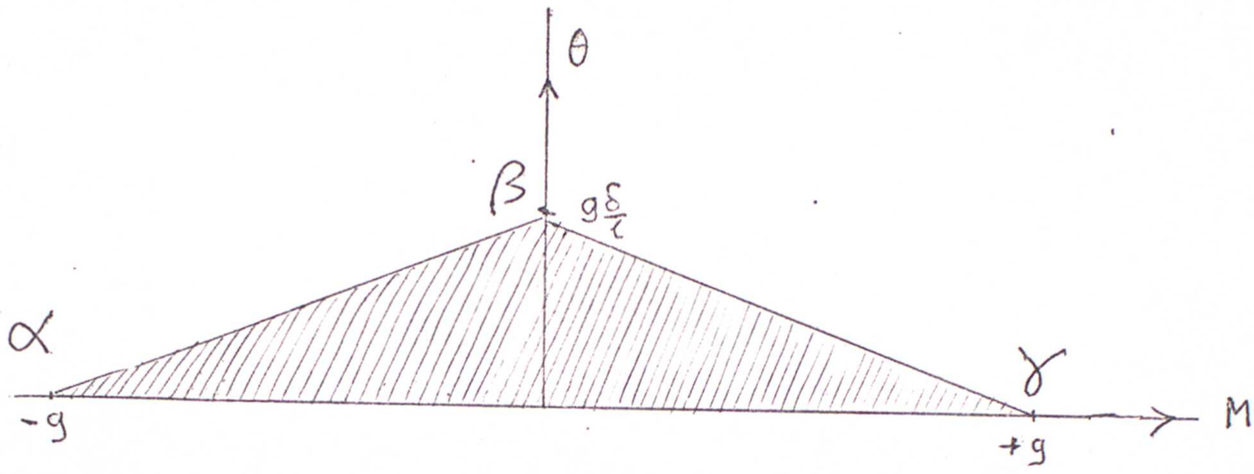


Figure 6

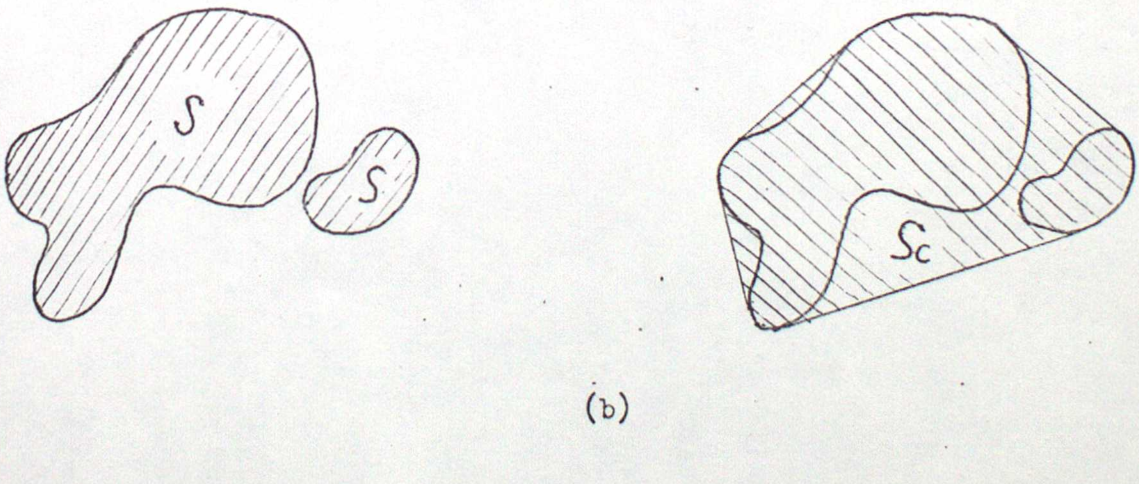


Figure 7

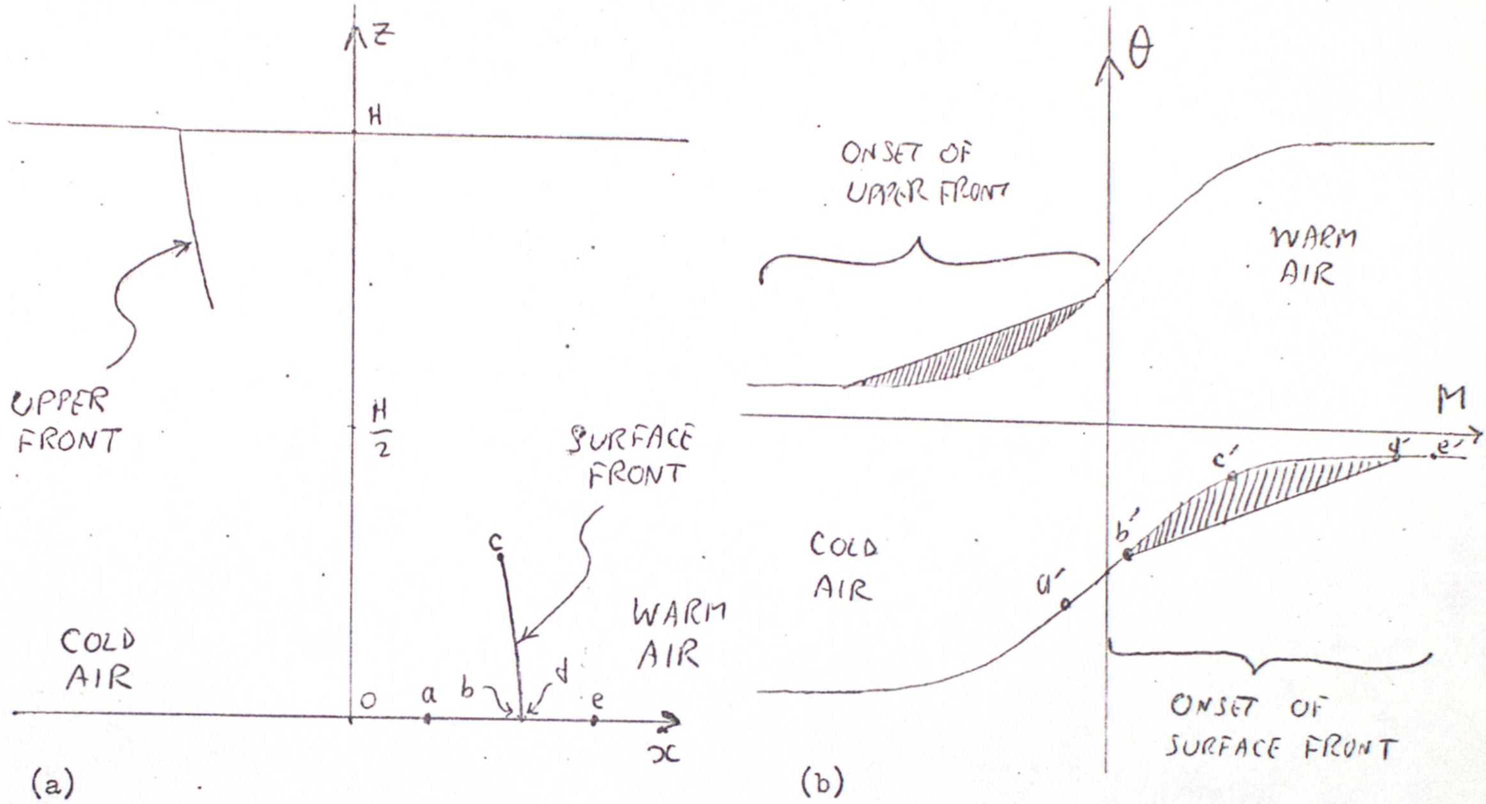


Figure 8

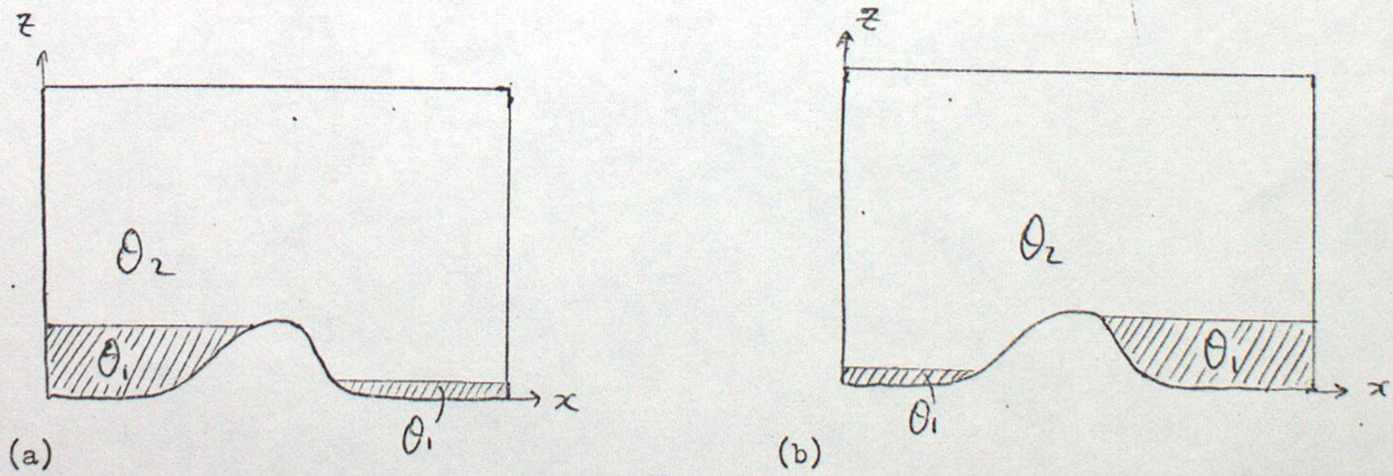


Figure 9