Bounds on translational and angular velocity components from first order derivatives of image flow

Muralidhara Subbarao
Department of Electrical Engineering, State University of New York
Stony Brook, NY 11794.

Abstract

A moving rigid object produces a moving image on the retina of an observer. It is shown that only the first order spatial derivatives of image motion are sufficient to determine (i) the maximum and minimum velocities of the object towards the observer, and (ii) the maximum and minimum angular velocities of the object along the direction of view. The second or higher order derivatives whose estimation is expensive and unreliable are not necessary. (The second order derivatives are necessary to determine the actual motion of the object; many researchers have worked on this problem.) These results are interpreted in the image domain in terms of three *differential invariants* of the image flow field: *divergence, curl,* and *shear magnitude.* In the world domain, the above results are interpreted in terms of the motion and local surface orientation of the object. In particular, the result that the *maximum velocity of approach* of an object can be determined from only the first order derivatives has a fundamental significance to both biological and machine vision systems. It implies that an organism (or a robot) can quickly respond to avoid collision with a moving object from only coarse information. This capability exists irrespective of the shape or motion of the object. The only restriction is that motion should be rigid.

1. Introduction

The relative motion of an observer with respect to an object produces a time-varying image on the observer's retina. This time-varying image contains valuable information about the three-dimensional (3D) shape and motion of the object. Recovering this information from the time-varying imagery is an important problem in computer vision.

The time-variation of an image can be represented by an *image velocity field* or an *image flow field.* An *image flow field* is a two-dimensional velocity field defined over the eye's retina (or image plane in the case of a camera). The velocity at any point is the instantaneous velocity of the image element at that point. Some authors refer to image flow as *optical flow.* Methods for the computation of image flow from time-varying images have been proposed by Horn and Schunck (1980), Hildreth (1983), Waxman and Wohn (1985), and others. The problem of three-dimensional interpretation of image flow has been addressed by many researchers (Longuet-Higgins and Prazdny, 1980; Longuet-Higgins, 1984; Kanatani, 1985; Waxman and Ullman, 1985; Subbarao and Waxman, 1986; Waxman, Kamgar-Parsi, and Subbarao, 1986; Subbarao, 1986a,b,c). In all these approaches, up to second order derivatives of image flow are used to recover the three-dimensional shape and motion of objects. The reliable estimation of the second order derivatives requires significant computation and very high quality images in terms of both spatial and gray level resolution. The human eye is very likely capable of exploiting the second order derivatives, but the present day machine vision systems are far from it (Adiv, 1985; Waxman and Wohn, 1985; Wohn and Waxman, 1985). Thus the requirements of high quality images and computational power have been major obstacles to using the already known theoretical results of image flow analysis in actual machine vision systems.

Obtaining a complete description of the shape and motion of an object may require a knowledge of the second or even higher order image flow derivatives, but some very useful information can be inferred from only up to the first order derivatives. For a given spatial and gray level resolution of the images, up to first order image flow derivatives can be recovered significantly more robustly than the second and higher order derivatives (Waxman and Wohn, 1985; Wohn and Waxman, 1985). In this paper we show that the first order flow derivatives are sufficient to determine the bounds on: (i) the velocity of approach of an object towards the observer, and, (ii) the angular velocity of the object along the direction of view. An interpretation of these two results are given in the image domain in terms of three *differential invariants* of the image flow field: *divergence, curl,* and *shear magnitude.* The boundary values of the translational and rotational velocities are related to these invariants by simple linear relations. The boundary values are also interpreted in the world domain in terms of the motion and local surface orientation of the object.

An object moving towards an observer could potentially collide and hurt the observer, or, in the case of a robot, damage the camera system. Therefore, in particular, the result that the *maximum velocity of approach* of an object can be determined from only the first order derivatives of image flow is of significance to both biological and machine vision systems. It implies that an organism (or a robot) can respond quickly to avoid collision with a moving object from only coarse information. This capability exists irrespective of the shape or motion of the object. The only restriction is that motion should be rigid.

For the special case where an observer is moving in a static environment, our results have an interesting consequence. (Examples of such a case are flying bees, birds, and helicopters.) In this case, by determining the bounds on the translational and angular velocities along some three mutually orthogonal viewing directions, bounds on the over all translational and rotational velocities of the observer can be determined from only first order image flow derivatives.

The results in this paper are potentially useful for collision avoidance by a robot in a dynamic environment and for robot navigation. Interestingly, biological vision systems have been found to be very quick in responding to approaching objects. This has been called the "looming effect" (Schiff, Caviness, and Gibson, 1962).

In the remaining part of this paper we derive the main results and give their interpretation in both the image domain and the world domain.

2. Camera geometry, notation, and image flow equations

A first approximation to the human eye is a pin-hole camera. For a global image flow analysis we suggest using a pin-hole camera with a spherical projection screen whose center is at the *pin-hole* or the *focus*. For this camera model, due to symmetry, the image flow analysis is identical at all points on the projection screen. However, here we do only a *local analysis* in a small field of view and in this field of view we consider the spherical screen to be approximated by a plane tangential to the spherical surface at the center of the field of view. The geometry of the screen is entirely a matter of convenience and does not affect our results. Note that there is a one to one correspondence between an image on a curved screen such as a spherical screen and an image on a planar screen. In our analysis using a planar projection screen, note that, the image flow being analyzed always corresponds to an object which is along a line normal to the image plane and passing through the focus. We call this line the *line of sight* or the *optical axis* or the *direction of view*.

The camera model is illustrated in Figure 1. The origin of a Cartesian coordinate system $OXYZ$ forms the *focus* and the Z-axis is aligned with the *optical axis*. The *image plane* is assumed to be at unit distance from the origin perpendicular to the optical axis. The image coordinate system oxy on the image plane has its origin at (0,0,1) and is aligned such that the x and y axes are, respectively, parallel to the X and Y axes.

Let the relative motion of the camera with respect to a rigid surface along the optical axis be described by translational velocity (V_X, V_Y, V_Z) and rotational velocity $(\Omega_X, \Omega_Y, \Omega_Z)$ around the focus. Also, let $Z = f(X, Y)$ represent the surface along the optical axis. The surface is assumed to be smooth. Let Z_X, Z_Y be the slopes of the surface at (X,Y)=(0,0) with respect to the X and Y axes respectively. Due to the relative motion of the camera with respect to the surface, a two-dimensional image flow is created by the perspective image on the image plane. At any point (x, y) on the image plane, let u, v be the components of image velocity along the x and y axes respectively. For the situation described here, Longuet-Higgins and Prazdny (1980) have derived the equations relating the derivatives of u, v at the image origin (up to second order) to the relative motion and shape of the surface. In these equations the translational velocity is always scaled by a quantity which cannot be determined. (This indeterminacy is due to the fact that absolute distance of objects cannot be determined using a monocular pin-hole camera. Therefore, a nearby object moving slowly and a distant object moving fast could both give rise identical image flows.) The scaling factor is usually chosen such that the distance of the surface along the optical axis is unity. Let the translational velocity scaled by this quantity be (V_x, V_y, V_z). At the image origin, let (u_0, v_0) be the image velocity and u_x, u_y, v_x, v_y be the partial derivatives of u, v with respect to the indicated subscripts x, y. The image velocity and its partial derivatives at the image origin describe the image flow in a small image region around the image origin. The following equations, originally derived by Longuet-Higgins and Prazdny (1980), represent the relation between the image flow and the shape and motion of the surface in a small field of view around the optical axis:

$$u_0 = - V_x - \Omega_Y \ , \quad v_0 = - V_y + \Omega_X, \tag{1a,b}$$

$$u_x = V_z + V_x \, Z_X \ , \qquad v_y = V_z + V_y \, Z_Y, \tag{1c,d}$$

$$u_y = \Omega_Z + V_x \, Z_Y \; v_x = - \Omega_Z + V_y \, Z_X. \tag{1e,f}$$

Above we have six equations in eight unknowns, hence an under constrained system of equations. We need more information to get a sufficiently constrained system of equations (e.g. see Longuet-Higgins and Prazdny, 1980; Waxman, Kamgar-Parsi, and Subbarao, 1986; Subbarao, 1986c). However we shall see that we can obtain bounds on the velocity of approach V_z and the angular velocity Ω_Z along the direction of view from these equations.

3. Bounds on the velocity of approach

First we state and prove a theorem which will be used later to establish bounds on the velocity of approach.

Theorem 1 : Suppose that translation parallel to the image plane is not zero and let r and θ be such that

$$V_x \equiv r \cos\theta \ \text{ and } \ V_y \equiv r \sin\theta \tag{2a,b}$$

$$\text{for } -\pi/2 < \theta \leq \pi/2 \ .$$

(Note: r is the signed magnitude of translation parallel to the image plane and θ is the direction of translation parallel to the image plane.) Then,

$$V_z = u_x \sin^2\theta + v_y \cos^2\theta - (u_y + v_x)\cos\theta \sin\theta. \quad (3)$$

Proof : From relations (1c-f), and (2a,b) we can get

$$u_y + v_x = r\cos\theta Z_Y + r\sin\theta Z_X \quad \text{and} \quad (4a)$$

$$u_x - v_y = r\cos\theta Z_X - r\sin\theta Z_Y . \quad (4b)$$

Solving for Z_X and Z_Y from above equations we get

$$Z_X = \frac{1}{r}\left\{(u_y + v_x)\sin\theta + (u_x - v_y)\cos\theta\right\} \quad \text{and} \quad (5a)$$

$$Z_Y = \frac{1}{r}\left\{(u_y + v_x)\cos\theta - (u_x - v_y)\sin\theta\right\} . \quad (5b)$$

Now, from relations (1c), (2a), and (5a) we can get

$$V_z = u_x - (u_y + v_x)\cos\theta\sin\theta - (u_x - v_y)\cos^2\theta . \quad (6)$$

Or, using the identity $\sin^2\theta + \cos^2\theta = 1$,

$$V_z = u_x(\sin^2\theta + \cos^2\theta) - (u_y + v_x)\cos\theta\sin\theta - (u_x - v_y)\cos^2\theta. \quad (7)$$

Relation (3) can be obtained from the above relation.

Notice that V_z, the velocity of approach along the direction of view, is given only in terms of θ. Therefore it can be determined if θ is known. Also it can be used to establish upper and lower limits on V_z.

Theorem 2 : The first order flow derivatives determine lower and upper bounds on the velocity of approach V_z of a surface along the line of sight. The bounds are

$$V_z^{(max/min)} = \frac{u_x + v_y}{2} \pm \frac{\sqrt{(u_y + v_x)^2 + (u_x - v_y)^2}}{2}. \quad (8)$$

Proof : By some trigonometric manipulation, expression (3) for V_z can be written as

$$V_z = \frac{u_x + v_y}{2} - \frac{u_y + v_x}{2}\sin 2\theta - \frac{u_x - v_y}{2}\cos 2\theta . \quad (9)$$

Differentiating the right hand side above and equating the resulting expression to zero we can show that the θs corresponding to the extrema of V_z are given by

$$\tan 2\theta = \frac{u_y + v_x}{u_x - v_y} . \quad (10)$$

From the above expression we have

$$\sin 2\theta = \frac{u_y + v_x}{\sqrt{(u_y + v_x)^2 + (u_x - v_y)^2}} \quad \text{and} \quad (11a)$$

$$\cos 2\theta = \frac{u_x - v_y}{\sqrt{(u_y + v_x)^2 + (u_x - v_y)^2}} \quad \text{for } 0 \le 2\theta \le 2\pi . \quad (11b)$$

Substituting for $\sin 2\theta$ and $\cos 2\theta$ from the above expressions in expression (9) we can get relation (8).

Note that all terms on the right hand side of relation (8) are only first order flow derivatives; no second or higher order derivatives are involved. Further, the above limits hold irrespective of the surface shape (except that the surface should be smooth because the image flow has been assumed to be differentiable).

4. Bounds on the angular velocity along the direction of view

Ω_Z is the angular velocity along the direction of view. By following in steps similar to the previous section, it can be shown that

$$\Omega_Z = u_y \sin^2\theta - v_x \cos^2\theta + (u_x - v_y)\cos\theta\sin\theta , \quad (12)$$

and

$$\Omega_Z^{(max/min)} = \frac{u_y - v_x}{2} \pm \frac{\sqrt{(u_y + v_x)^2 + (u_x - v_y)^2}}{2}. \quad (13)$$

5. Interpretation of the bounds in the image domain

In order to interpret the bounds on V_z and Ω_Z we make the following observation. To a first order, the image velocity field in a small field of view around the direction of view can be described by

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (14)$$

The above expression represents an *affine transformation.* In this expression, the vector $[u_0, v_0]^T$ gives the *pure translation* of the image region at the image origin; the 2×2 tensor on the right hand side is the *velocity gradient tensor.* This tensor can be expressed uniquely as the sum of a *symmetric tensor* and an *anti-symmetric tensor* as below:

$$\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \begin{bmatrix} u_x & (v_x + u_y)/2 \\ (v_x + u_y)/2 & v_y \end{bmatrix} + \begin{bmatrix} 0 & (u_y - v_x)/2 \\ -(u_y - v_x)/2 & 0 \end{bmatrix} . \quad (15)$$

In Fluid Mechanics literature (e.g.: Aris, 1962), the symmetric tensor of a velocity gradient tensor is called the *deformation* or *rate of strain* tensor and the anti-symmetric tensor is called the *spin tensor.* These tensors have nice physical interpretations. We will borrow these well known ideas from Fluid Mechanics to interpret our results. Such an interpretation of image flow has already been described by many others in the computer vision area (Koenderink and Van Doorn, 1975, 1976; Waxman and Ullman, 1985; Kanatani, 1986).

The independent parameter $u_y - v_x$ of the spin tensor is called the *spin* or *vorticity.* It is also the *negative curl* of the image velocity field at the image origin, i.e.

$$-\text{curl} = u_y - v_x . \quad (16)$$

This can be easily verified from relation (14). It gives the *rigid body rotation* of the image neighborhood at the image origin. By setting all terms except the curl term to zero, i.e.

$$u_0 = v_0 = (v_x + u_y) = u_x = v_y = 0, \quad (17)$$

we can obtain the image flow field corresponding to this term. The term results in a purely *rotational flow field.*

The deformation tensor gives the deformation of the image neighborhood at the image origin. We can interpret this tensor in terms of its eigen values. The two eigen values of this tensor are in fact $V_z^{\max}$, $V_z^{\min}$, given by relation (8). The sum of the eigen values (which is also the trace of the original tensor) is the *divergence* of the image velocity field at the image origin, i.e.

$$\text{divergence} = u_x + v_y \ . \tag{18}$$

This can be easily verified from relation (14). This quantity gives the *isotropic expansion* or contraction of the image neighborhood at the image origin. The image flow corresponding to the divergence term is obtained by setting other terms to zero, i.e.

$$u_0 = v_0 = u_y = v_x = (u_x - v_y) = 0 \ . \tag{19}$$

The result is a *purely divergent flow.*

The difference of the two eignen values of the deformation tensor is the *magnitude of pure shear* of the image neighborhood at the image origin, i.e.,

$$\text{Shear magnitude} = \sqrt{(u_y + v_x)^2 + (u_x - v_y)^2} \ . \tag{20}$$

The image neighborhood undergoes a contraction along one direction and an expansion orthogonal to it under constant area. The directions of contraction and expansion are aligned with the two eigen vectors of the deformation tensor. The image flow corresponding to a pure shear transformation is obtained by setting all but the shear terms to zero, i.e.,

$$u_0 = v_0 = u_x + v_y = u_y - v_x = 0 \ . \tag{21}$$

An example of a pure shear flow is shown in Figure 2.

In summary, a small circular image element at the image origin translates rigidly with velocity $[u_0, v_0]^T$, rotates as a rigid area with spin $u_y - v_x$, dilates according to the sum of the eigen values of the deformation tensor, and undergoes a stretch and compression at constant area according to the difference of the eigen values of the deformation tensor (along mutually orthogonal axes aligned with the eigen vectors) (Koenderink and Van Doorn, 1975, 1976; Waxman and Wohn, 1986).

In view of our above discussion and equations (16,18,20), equations (8,13) which give bounds on V_z and Ω_Z can be expressed as below.

Maximum/Minimum approach velocity

$$= \frac{1}{2}\left(\text{Divergence} \pm \text{Shear magnitude}\right) . \tag{22}$$

$$\left\{\begin{array}{l}\text{Maximum/Minimum angular velocity}\\ \text{around the viewing direction}\end{array}\right\}$$

$$= \frac{1}{2}\left(-\text{Curl} \pm \text{Shear magnitude}\right) . \tag{23}$$

The quantities: divergence, curl, and shear magnitude are all *invariant* with respect to the orientation of the image axes. Their values are unaffected by a rotation of the image coordinate system. This can be easily shown by considering how the image flow derivatives u_x, u_y, v_x, v_y are transformed by a rotation of the image coordinate system (e.g. see Kanatani, 1986). Hence they are called *differential invariants* of image flow.

6. Interpretation in the world domain

Let us now interpret what the bounds mean in the world domain. For this sake we introduce two vectors, $\boldsymbol{r}$ which is the direction of translation parallel to the image plane, and $\boldsymbol{p}$ which is the gradient of the object's surface with respect to the image plane. More specifically, if $\boldsymbol{i}$, $\boldsymbol{j}$ are unit vectors along the X, Y axes respectively, then, let

$$\boldsymbol{r} = \boldsymbol{i}V_x + \boldsymbol{j}V_y \ , \quad \text{and} \quad \boldsymbol{p} = \boldsymbol{i}Z_X + \boldsymbol{j}Z_Y \ . \tag{24a,b}$$

Now, from equations (1c,d,18,24a,b) we can show that

$$\text{divergence} = 2V_z + \boldsymbol{r}\cdot\boldsymbol{p} \ . \tag{25a}$$

Let $\boldsymbol{k}$ be a unit vector along the Z axis. Then, from equations (1e,f,16,24a,b) we can show that

$$-\text{curl}\ \boldsymbol{k} = 2\Omega_Z \boldsymbol{k} + \boldsymbol{r}\times\boldsymbol{p} \ . \tag{25b}$$

Also, from equations (1c-f,20,24a,b) we can show that

$$\text{Shear magnitude} = |\boldsymbol{r}|\,|\boldsymbol{p}| \ . \tag{25c}$$

The above relations (25a-c) show how the differential invariants of image flow are related to the three-dimensional motion and surface orientation. Some of the terms in these relations are in agreement with our intuition, for example the appearance of V_z in divergence and Ω_Z in curl. Now, from equations (22,23,25a-c) we can show that

$$V_z^{(max/min)} = V_z + \frac{1}{2}\left(\boldsymbol{r}\cdot\boldsymbol{p} \pm |\boldsymbol{r}|\,|\boldsymbol{p}|\right) \quad \text{and} \tag{26a}$$

$$\boldsymbol{k}\ \Omega_Z^{(max/min)} = \boldsymbol{k}\ \Omega_Z + \frac{1}{2}\left(\boldsymbol{r}\times\boldsymbol{p} \pm \boldsymbol{k}\ |\boldsymbol{r}|\,|\boldsymbol{p}|\right) \ . \tag{26b}$$

The above relations show how the bounds are related to the translation parallel to the image plane $\boldsymbol{r}$ and the surface gradient with respect to the image plane. We are not able to give a straightforward physical interpretation of the above two relations, but they seem to have a pleasing form. We believe that an interpretation of these equations is related to the discussion in Koenderink and Van Doorn (1975) about the different types of image flows generated depending on the eigen values of the velocity gradient tensor.

7. Conclusion

We have shown that using only the first order derivatives of the image flow of an object, a monocular observer can determine the bounds on (i) the translational velocity of the object towards the observer, and (ii) the angular velocity of the object in the direction of its position with respect to the observer. These bounds are related to the three first order differential invariants: divergence, curl, and shear magnitude of the image flow by simple linear relations. The above results are potentially useful in collision avoidance with moving objects by robot systems and also in autonomous navigation of vehicles. The results also throw some light on the computational aspects of the *looming effect* phenomenon observed in biological organisms.

Acknowledgement: I thank Dr. Allen Waxman for his helpful comments on this work.

8. References

[1] Adiv, G. 1985 (July). Determining 3-D motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7, pp. 384-401.

[2] Aris, R. 1962. *Vectors, Tensors, and the Basic Equations of Fluid Mechanics.* Englewood Cliffs: Prentice-Hall.

[3] Hildreth, E.C. 1983 (August). The Measurement of Visual Motion. Ph.D. Thesis, M.I.T.

[4] Horn, B.K.P., and B.G. Schunck. 1981. Determining optical flow. *Artificial Intelligence,* 17, pp. 185-204.

[5] Kanatani, K. 1985. Structure from motion without correspondence: general principle. *Proceedings of 9th International Joint Conference on Artificial Intelligence*, pp. 886-888.

[6] Kanatani, K. 1986. Group theoretical methods in image understanding. CAR-TR-214, Center for Automation Research, University of Maryland.

[7] Koenderink, J.J. and van Doorn, A.J. 1975. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. Optica Acta 22, pp. 773-791.

[8] Koenderink, J.J. and van Doorn, A.J. 1976a. Local structure of movement parallax of the plane. J. Opt. Soc. Amer. 66, pp. 717-723.

[9] Longuet-Higgins, H.C., and K. Prazdny. 1980. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, B 208, pp. 385-397.

[10] Longuet-Higgins, H.C. 1984. The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London*, B 223, pp. 165-175.

[11] Subbarao, M. 1986a (April). Interpretation of image motion fields: rigid curved surfaces in motion. CAR-TR-199, Center for Automation Research, University of Maryland.

[12] Subbarao, M. 1986b (May). Interpretation of image motion fields: a spatio-temporal approach. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, pp. 157-166.

[13] Subbarao, M. 1986c. Interpretation of visual motion: a computational study. Ph.D. Thesis, Department of Computer Science, University of Maryland.

[14] Subbarao, M., and A.M. Waxman. 1986. Closed-form solutions to image flow equations for planar surfaces in motion. *Computer Vision, Graphics, and Image Processing,* **36,** pp. 208-228.

[15] Waxman, A.M., B. Kamgar-Parsi, and M. Subbarao. 1986 (February). Closed-form solutions to image flow equations for 3-D structure and motion. CAR-TR-190, Center for Automation Research, University of Maryland.

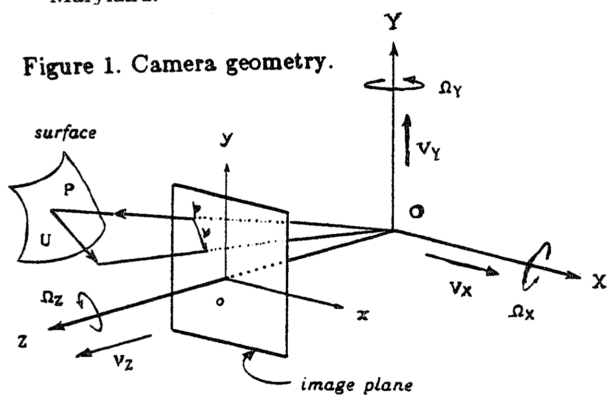
[16] Waxman, A.M., and S. Ullman. 1985. Surface structure and three-dimensional motion from image flow kinematics. *International Journal of Robotics Research* 4 (3), pp. 72-94.

[17] Waxman, A.M. and K. Wohn. 1985. Contour evolution, neighborhood deformation and global image flow: planar surfaces in motion. *International Journal of Robotics Research*, 4 (3), pp. 95-108.

[18] Waxman, A.M. and K. Wohn. 1986. Image flow theory: a framework for 3-D inference from time-varying imagery. A chapter in *Advances in computer vision,* Ed. C. Brown, Erlbaum publishers.

[19] Wohn, K. and A.M. Waxman. 1985 (July). Contour evolution, neighborhood deformation and local image flow: Curved surfaces in motion. CAR-TR-134, Center for Automation Research, University of Maryland.

Figure 1. Camera geometry.



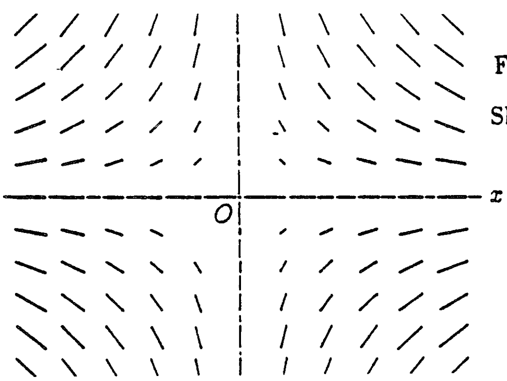


Figure 2. Shear flow.