Journal of Geodesy manuscript No.
(will be inserted by the editor)

The orthometric height and the holonomity problem

Fernando Sansò[1], **Petr Vaníček**[2]

[1] DIIAR, Politecnico di Milano, I-20133, Italy, Phone 0039-031-3327518, Fax 0039-031-3227519 e-mail: fernando.sanso@polimi.it, web: http://www.polimi.it.

[2] Department of Geodesy and Geomatic Engineering, University of New Brunswick, Fredericton, NB, E3B 5A3, Canada, Phone 506-453-5144, Fax 506-453- 4943, e-mail: vanicek@unb.ca, web: http://gge.unb.ca.

Received: 03/03/2005 / Accepted: 31/10/2005

Abstract. When height networks are being adjusted, many geodesists advocate the approach where the adjustment should be done by using geopotential numbers rather than the orthometric or normal heights used in practice. This is based on a conviction that neither orthometric nor normal heights can be used for the adjustment because these height systems are not holonomic, meaning - among other things - that height increments (orthometric or normal) when summed around a closed loop do not sum up to zero. If this were the case, then the two height systems could not be used in the adjustment; the non-zero loop closure would violate the basic, usually unspoken, assumption behind the adjustment, namely that the model claiming that height differences are observable is correct. In this paper, we prove in several different ways that orthometric and normal heights are theoretically just as holonomic as the geopotential numbers are, when they are obtained from levelled height differences using actual gravity values. This disposes of the argument that geopotential numbers should be used in the adjustment. Both orthometric and normal heights are equally qualified to be used in the adjustment directly.

Send offprint requests to: Fernando Sansò

Key words. Heights, holonomity, gravity.

1 Motivation

The question whether the orthometric height (H) is or is not a proper coordinate, more precisely whether it is holonomic or anholonomic, has been sometimes debated in geodesy; this has been related to a thorough analysis of the gravity field in geometric terms, which has been performed in a number of basic works like Hotine (1969), Marussi (1985), Grafarend (1975, 1995). However, although the answer to the above question in the affirmative sense is very clear and undisputable, it is from time to time raised again. This is due to a misinterpretation of the equation

$$dW = -g\, dH \tag{1}$$

which, without any further specification, is definitely wrong.

As a matter of fact, Eq. (1) is typically justified by a local argument as follows. Let us take a point $\overline{P}$ in space and the equipotential surface $\{W_P = W_{\overline{P}}\}$ through it; in $\overline{P}$ we can define a local Cartesian triad (a Cartan frame) with the Z axis up (i.e., contrary to the gravity vector $\underline{g} = \nabla W)$ and the X and Y axes towards the astronomical East and North directions, respectively. If we analyze the variation of W around $\overline{P}$, we see that it varies only when we move along the Z axis because in the other two directions we are tangent to the equipotential surface. Since <u>along the vertical</u>, by definition,

$$dH = dZ \tag{2}$$

Eq. (1) holds <u>in that direction</u>, because $g(\overline{P}) = |\nabla W|$ and g points downward. The logical error here is that, although Eq. (1) is true along the vertical, implying that, in a coordinate system for which the plumbline is also a coordinate line,

$$-g(\overline{P}) = \frac{\partial W}{\partial H}(\overline{P})\ , \tag{3}$$

the same is not true when we move in the horizontal because then, as we shall see, dH is not zero!

In addition, when we claim that equations like Eqs. (1) or (3) are valid, we tacitly assume that H is just a function of the point $\overline{P}$, i.e., that it is a holonomic coordinate, so we cannot use these equations to judge on this property, but rather we have to first clarify independently that it is a holonomic coordinate and then state whether Eqs. (1) or (3) are correct.

The understanding of this point in geodesy is so fundamental that the authors deem it useful to write this article, where what can be found in classical textbooks (Heiskanen and Moritz (1967)) is highlighted and strengthened, pointing out where misunderstandings could have originated.

On the other hand, whether Eq. (1) is wrong or not depends on the examined gravity field, because $H_{(P)}$, the curved distance computed along the plumbline of the point P from a reference equipotential (the geoid), is tailored onto the particular field considered.

For instance, it is obvious that Eq. (1) holds true for a purely spherical (non-rotating) symmetrical model of the Earth, because indeed the monopole potential $(\mu = GM)$,

$$u = \mu/r \tag{4}$$

has spherical equipotential surfaces and purely radial gravity, so that $dH = dr$ and

$$du = -\frac{\mu}{r^2}dr = -g\; dH\; . \tag{5}$$

The question then arises whether the spherical potential is the only one for which Eq. (1) holds true. Such a question is less trivial than what appears at a first sight. However, if we remain in a hypothetical model of a non-rotating Earth, a general answer becomes straightforward by exploiting the so-called gravity space formalism.

2 A purely mathematical argument

The fact that H is a true, holonomic, coordinate is just an application of the definition of a coordinate system and it depends purely on whether the hypotheses implicit in the application of the fundamental integration lemma, can be accepted or not.

Therefore, let us start with the basic definition.

Definition 1 A coordinate system in a domain Ω of R^3 is a system of three real functions of the point

$$P, (X_1(P), X_2(P), X_3(P)),$$

all of them defined in Ω in such a way that there exists a one-to-one correspondence (cf. (Marussi, 1985), (Grafarend, 1975))

$$P \Leftrightarrow (X_1, X_2, X_3) \tag{6}$$

.

Typically, we require some regularity property from a coordinate system, such as continuity or even differentiability. For the moment we will assume that such properties are satisfied for all systems discussed in the sequel, possibly except for some singular points – think of the origin in a spherical coordinate system.

Remark 1 An anholonomic coordinate (H^*), on the other hand, is just a number assigned to a point P and to a path connecting P to some origin P_0. Typically, this is done by defining a differential form

$$\delta H^*(P, d\ell) \equiv \underline{\eta}(P) \cdot d\underline{\ell} \tag{7}$$

expressing only an increment of H^* when we move from the point P to $P + \delta P$, such that $\delta P = d\underline{\ell}$, and then by stating that, given a curve Γ connecting P_0 and P and assigning an arbitrary but fixed value to $H^*(P_0)$ (e.g., $H^*(P_0) = 0$), we put

$$\begin{aligned} H^*(P, \Gamma) &= H^*(P_0) + \int_\Gamma \delta H^*(P, s\underline{\ell}) = \\ &= H^*(P_0) + \int_\Gamma \underline{\eta}(P) \cdot d\underline{\ell} \; . \end{aligned} \tag{8}$$

In this sense, an anholonomic coordinate is not a coordinate as long as Eq. (8) depends on the path Γ. In other words, if we wish H^* to be independent of Γ and therefore to be a true coordinate, the integral of δH^* over any closed loop has to be equal to zero

$$\oint_\Gamma \delta H^*(P, d\underline{\ell}) = \int_\Gamma \underline{\eta}(\sigma) \cdot d\underline{\ell} \equiv 0 \; \forall \Gamma \; . \tag{9}$$

One the most famous theorems of vector calculus is that of Stokes (Kaplan (1991)). It claims:

Theorem 1 (GG Stokes) δH^* is an exact differential, i.e., Eq. (9) holds for a smooth field $\underline{\eta}$ in Ω if and only if

$$\text{rot}\, \underline{\eta} = \nabla \wedge \underline{\eta} \equiv 0 \tag{10}$$

and the domain Ω is arcwise simply connected, i.e., such that any loop Γ in Ω can be shrunk to a point by a continuous deformation, without getting out of Ω.

Just as a remark, we note that when H^* is known to be a coordinate, then Eq. (10) is automatically satisfied since in this case

$$\underline{\eta} = \nabla H^* \tag{11}$$

and the rotor of a gradient is identically equal to zero.

Already on that purely logical ground we are able to draw the first conclusion, namely that the orthometric height (H), as a pure function of a point P in space, is indeed a proper, holonomic, coordinate, under suitable regularity of the gravity field.

Conclusion 1 Assume that the gravity field of a spherical body is given in terms of a potential W obtained by adding to Newtonian potential, generated by a bounded mass-density, a centrifugal potential $\frac{1}{2}\omega^2(x^2+y^2)$ (i.e., z is taken along the axis of rotation); then it is known (Miranda (1970)) that both the function W and its gradient $\underline{g} = \nabla W$ are continuous throughout the whole R^3 and that at points far away from the barycenter, the following asymptotic expressions hold

$$\begin{aligned} &W \sim \frac{\mu}{r} + \frac{1}{2}\omega^2(x^2+y^2) + 0\left(\frac{1}{r^2}\right) , && (12)\\ &\underline{g} = \nabla W \sim -\frac{\mu \underline{r}}{r^3} + \omega^2(x\underline{e}_x + y\underline{e}_y) + 0\left(\frac{1}{r^3}\right) . \end{aligned}$$

In Eqs. (12) and (13), the terms $0\left(\frac{1}{r^2}\right)$ and $0\left(\frac{1}{r^3}\right)$ became respectively $0\left(\frac{1}{r^3}\right)$ and $0\left(\frac{1}{r^4}\right)$ as soon as we stipulate to place the origin of the coordinates at the barycenter of the masses generating the field. Assume further that ω is small enough to allow the existence of a system of equipotential surfaces $\{P; W(P) = \text{const}\}$ that are smooth (at least with a continuous normal) and closed, up to some surface enclosing all the masses (remember that equipotential surfaces of a rotating body are quite complicated and include pieces of unbounded surfaces, so we refer here to the bounded closed surfaces surrounding the origin). Let us call Ω the bounded region between two equipotential surfaces

$$\begin{aligned} &S_0 \equiv \{W = W_0\} && (13)\\ &S_e \equiv \{W = W_e\} \ , \ W_e \leq W_0 \ , \end{aligned}$$

such that S_0 is taken as a reference surface (the geoid), which is close to the surface of the masses (partially inside), while S_e is totally external to the masses, so that we are sure that there is one and only one equipotential surface that is passing through each point in Ω. Assume finally that the plumblines, i.e., the trajectories orthogonal to the equipotentials in the region Ω, are a regular family of curves, defined in parametric form by the equations

$$\underline{\dot{x}} = -\nabla W = -\underline{g} \;, \tag{14}$$

such that there is one and only one line passing through each point P of Ω.

Indeed this is just one of the possible choices that are studied and described in literature (e.g., Grafarend (1975)); all of them have to be equivalent, as for the description of the family of the plumblines.

Then, given any point P in Ω, we have the corresponding (unique) plumbline through it and we also have only one point P_0 on S_0 through which the same plumbline passes. Accordingly, given P we can compute the unique length of the arc of plumbline

$$H = \int_{P_0}^{P} d\ell = \int_{P_0}^{P} |\underline{\dot{x}}| dt = \int_{P_0}^{P} g \; dt \;. \tag{15}$$

Therefore, H is a function of P only, for a given gravity field, *and is thus a true coordinate*.

A formula totally equivalent to Eq. (15) is

$$H_P = -\int_{P_0}^{P} \frac{dW}{g} \;, \tag{16}$$

which, in order to avoid misunderstandings, has to be clearly interpreted as an *integral along the single plumbline that passes through* P. The equivalence between Eqs. (15) and (16) comes from observing that, according to Eq. (14),

$$-dW = -\nabla W \cdot \underline{\dot{x}} dt = |\underline{\dot{x}}|^2 dt = g^2 dt \;, \tag{17}$$

which substituted in Eq. (16) gives Eq. (15).

Remark 2 It is important to underline that Eq. (16) *does not imply* a relation like

$$dH = -\frac{dW}{g} \;, \tag{18}$$

as we shall prove more extensively in the next Section by explicitly computing the differential of the function H. Here, however, we provide a short proof by a reduction *ad absurdum*.

Assume Eq. (18) to be true, then the equipotential surfaces should also coincide with surfaces of constant orthometric height, because

$$dW = 0 \Rightarrow dH = 0 \ . \tag{19}$$

If one assumes only that W decreases along a plumbline, which is a by-product of the definition in Eq. (14), from Eq. (19) we could conclude that

$$W = W(H) \tag{20}$$

and therefore

$$g = -\frac{dW}{dH} \tag{21}$$

is also a function of H only. In this case, g should be constant on equipotential surfaces, which runs contrary to experimental truth (observation).

A good visual perception of this fact comes from inspecting the free-air gravity anomaly map of the oceans (Andersen and Knudsen, 1998). Therefore, Eq. (18) cannot be true for the configuration of plumblines of the actual gravity field.

3 The differential of the orthometric height

The purpose of this Section is to derive the very classical formula (cf. Heiskanen and Moritz (1967), §4.4)

$$dH = \delta L + \delta H^* \tag{22}$$

where dH is the differential of the function $H = H(P)$ when P undergoes a displacement $dP = d\underline{\ell}$, where δL is the leveling increment, the observable quantity, and δH^* is the so-called orthometric correction, denoted by the symbol OC. The reason why we want to redo such a derivation is that we would like to highlight two interpretations of δH^* that are known (cf. Grafarend (1997)) but not current, though very expressive of the meaning of such a quantity from both the geometrical and the mechanical point of view.

Remark 3 In pursuing the above aim, we will verify at the same time that H_P is a differentiable function, for realistic mass distributions, namely distributions with a bounded density ρ_P. The basic statement here is that when ρ_P is bounded, with a number of discontinuity surfaces, then W_P and $\underline{g}_P = \nabla W$ are continuous throughout the space, while the matrix of second derivatives of $W_P, M = \left[\frac{\partial^2 W}{\partial X_i \partial X_k}\right]$, can present at most a logarithmic behaviour at the surfaces of discontinuity of ρ, as we know from elementary computations of the potential of a uniform prism (e.g., MacMillan (1958)).

Figure 1 near here

In order to proceed, we first carefully inspect Fig. 1. First, we see that

$$dH = H_{P+dP} - H_P = (H_{P+dP} - H_{P^*}) + \\ +(H_{P^*} - H_P) = \delta L + \delta H^* \tag{23}$$

where

$$\delta L = H_{P+dP} - H_{P^*} = \underline{e}_H(P) \cdot d\underline{\ell} = \\ = -\frac{\underline{g}(P)}{g_P} \cdot d\underline{\ell} = -\frac{dW_P}{g_P} \tag{24}$$

is the leveling increment, while $\delta H^* = H_{P^*} - H_P$ is the orthometric correction amounting to the variation of the orthometric height when we move P on the equipotential $\{W = W_P\}$, to reach P^*.

Note that, for the sake of readability, the height H^{P^*}, i.e. the arclength $P^* P_0 + \delta P_0$, is not drawn in Fig. 1. Also, please note that here dW_p is not the increment of W_p along a particular plumbline, but just its spatial differential from P to $P + \delta P$.

Conclusion 2 If we combine Eqs. (24) and (23), we see that

$$dH_P = -\frac{dW_P}{g_p} + \delta H^*_{(PP^*)} , \tag{25}$$

which proves that Eq. (18) is a *wrong equation*, unless

$$\delta H^*_{(PP^*)} \equiv 0 \quad (\forall \delta P) . \tag{26}$$

Indeed, if $\delta H^* \equiv 0$ in every direction, then the orthometric heights of points on $\{W = W_P\}$ are all constant, i.e. the potential W_P is function of H_P and we return to the argument of Remark 2.

Now, to better understand the geometric significance of δH^*, we can use the relation

$$H_P = h_P - N \ , \ H_{P^*} = h_{P^*} - N' \tag{27}$$

to find

$$\frac{dH^*}{d\ell_0} = \left(\frac{h_{P^*} - h_P}{d\ell_0}\right) - \left(\frac{N' - N}{d\ell_0}\right) \tag{28}$$

which shows that, if E is the reference ellipsoid,

$$\begin{aligned}\frac{dH^*}{d\ell_0} = \varepsilon_P - \varepsilon_0 = & \text{[inclination of } \{W = W_P\} \text{ on } E] \\ & - \text{[inclination of } G \text{ on } E] \ ;\end{aligned} \tag{29}$$

namely $\frac{dH^*}{d\ell_0}$ is just the inclination, in the direction $d\underline{\ell}_0$, of the equipotential surface through P with respect to the equipotential through P_0, i.e., the geoid G.

Observe that ε_p is just the Helmert deflection of the vertical in the direction of $d\underline{\ell}_0$, with opposite sign, at the point P, while ε_0 is the corresponding quantity at the geoid. Therefore, recalling that

$$\begin{aligned}&-\varepsilon_P d\ell_0 = \nabla_0 \left(\tfrac{T_P}{\gamma_P}\right) \cdot d\underline{\ell}_0 \ , \ +\varepsilon_0 d\ell_0 = \nabla_0 \left(\tfrac{T_{P_0}}{\gamma_{P_0}}\right) \cdot d\underline{\ell}_0 \ \ , \\ &\qquad\qquad (\nabla_0 = \text{ horizontal gradient})\end{aligned} \tag{30}$$

from Eq. (28) one can express δH^* in terms of the anomalous potential T. This will be further worked out in Remark 5.

Here, we would like to show that yet another way of expressing δH^* is (cf. Fig. 1)

$$\begin{aligned}\delta H^* = & \int_{P_0+\delta P_0}^{P+\delta P} dH_{Q+\delta Q} - \int_{P_0}^{P} dH_Q = \\ = & \int_{P_0}^{P} \left(\frac{dW_Q}{g_{Q+\delta Q}} - \frac{dW_Q}{g_Q}\right) \ ,\end{aligned} \tag{31}$$

because, by definition, $dW_{Q+\delta Q} = dW_Q$ (see Fig. 1). On the other hand, we know that (cf. Heiskanen and Moritz (1967) Ch. 2 §3)

$$dg_Q = \nabla_0 g_Q \cdot d\underline{\ell}_0 = g_Q K_Q \underline{n}(Q) \cdot d\underline{\ell} \tag{32}$$

where $\underline{n}(Q)$ is the principal normal of the plumbline (which therefore is tangent to the equipotential surface) and K_Q its principal curvature. In Eq. (32), we have taken into account that

$\underline{n} \cdot d\underline{\ell}_0 = \underline{n} \cdot d\underline{\ell}$, to a degree of relative accuracy reaching the level of $\sim 5 \cdot 10^{-8}$ when the deflection of the vertical is 1 arc-minute. By using Eq. (32), we derive

$$\begin{aligned}\delta H^* &= \int_{P_0}^{P} \frac{dW_Q}{g_Q^2} g_Q K_Q \underline{n}_Q \cdot d\underline{\ell} = \\ &= \left(- \int_{P_0}^{P} dH_Q K_Q \underline{n}_Q \right) \cdot d\underline{\ell} \,.\end{aligned} \tag{33}$$

In Eq. (33), we immediately read that if one has to have $\delta H^* \equiv 0, \ \forall P$ and $\forall d\underline{\ell}$, i.e., if Eq. (18) has to be true, one must conclude that

$$K_Q \equiv 0 \,, \tag{34}$$

i.e., the plumblines have to be straight lines because they must have zero curvature.

Remark 4 By combining Eqs. (22), (32) and (33), we can see that dH is a first-order differential in the displacement $d\underline{\ell}$, where the coefficients are continuous functions of P. As a matter of fact, by using the statement of Remark 3, we see that δL involves only the vector $\underline{g}$, which is continuous, while δH^* involves the second derivatives of W, which have at most logarithmic discontinuities; these however, once integrated along the plumbline yield continuous functions, as we have maintained in Remark 3.

Remark 5 Here we would like to elaborate a little on the differential expression of the orthometric correction implicit in Eq. (28) and (30) showing its equivalence to a more ordinary expression (cf. Heiskanen and Moritz (1967) Ch. 4, §4).

We start from

$$\delta H^* = \nabla_0 \left[\frac{T_P}{\gamma_P} - \frac{T_{P_0}}{\gamma_{P_0}} \right] \cdot d\underline{\ell}_0 \tag{35}$$

and note that

$$\begin{aligned}F(P) = \frac{T_P}{\gamma_P} - \frac{T_{P_0}}{\gamma_{P_0}} &= \int_{P_0}^{P} \left(\frac{\partial}{\partial H} \frac{T_Q}{\gamma_Q} \right) dH \cong \\ &\cong \int_{P_0'}^{P} \left(\frac{\partial}{\partial H} \frac{T_Q}{\gamma_Q} \right) dh \,,\end{aligned} \tag{36}$$

the last step being justified by the fact that the small change of integration path is applied to a quantity already small at the first order.

In fact, $\frac{dH - dh}{dh}$ is of the order of $\frac{1}{2}\varepsilon^2$, while the integrand has the order of the ratio of a gravity anomaly with normal gravity.

On the other hand, we know from the fundamental equation of physical geodesy (cf. Heiskanen and Moritz (1967) Ch. 2, §13) that

$$\frac{\partial}{\partial h}\frac{T}{\gamma} = -\frac{\Delta g}{\gamma} \; ; \tag{37}$$

hence Eq. (36) reads

$$F(P) = -\int_{P_0'}^{P} \frac{\Delta g_Q}{\gamma_Q} dh \; . \tag{38}$$

We use this relation in the following identity

$$\begin{aligned} \delta H^* &= \nabla_0 F(P) \cdot d\underline{\ell} = \left[\nabla F(P) - \frac{\partial F(P)}{\partial h}\underline{\nu}_P\right] \cdot d\underline{\ell} = \\ &= dF(P) - \frac{\partial F}{\partial h}\delta L = -d\left(\int_{P_0'}^{P} \frac{\Delta g_Q}{\gamma_Q} dh\right) + \frac{\Delta g_P}{\gamma_P}\delta L \end{aligned} \tag{39}$$

thus arriving to the sought-after expression of δH^* as a function of T.

Moreover, we see that by integrating Eq. (39) along a path from a point A to a point B on the Earth's surface, we get the more common expression

$$\begin{aligned} OC(AB) = \int_A^B \delta H^* &= -\int_{B_0'}^{B} \frac{\Delta g}{\gamma} dh + \int_{A_0'}^{A} \frac{\Delta g}{\gamma} dh + \\ &+ \int_A^B \frac{\Delta g}{\gamma} \delta L \; , \end{aligned} \tag{40}$$

where the first two integrals run along ellipsoidal normals (or along plumblines, according to our level of approximation), while the third integral has to be performed along the leveling path on the Earth's surface.

Conclusion 3 Although the orthometric height is a proper coordinate, with a continuous differential, yet it is not related in an elementary form to observable quantities, like the leveling increment δL. The orthomeric correction (δH^*), necessary to relate δL to the differential dH, is intimately related to the knowledge of the mass-density between the actual topographic surface and the geoid, and therefore to the modelling of the gravity field therein. In this sense, H has to be considered as a derived quantity rather than a primary coordinate for the description of geodetic height systems.

4 Are there gravity fields with parallel equipotential surfaces?

Up to now, we have been able to understand that Eq. (18) is not correct for the actual gravity field of the Earth, because otherwise W and g should be constant on the same surfaces, which is not true in our physical reality.

In addition, we have seen that if Eq. (18) would hold, then the plumblines should be straight lines and the level surfaces should be geometrically parallel to one another, because H should also be constant on these surfaces. We have already mentioned that there is at least one physical situation in which such conditions are verified; a completely spherically layered mass distribution in the case of a *non-rotating body*.

In this case, as the body itself has to be a sphere of some radius R and we know that

$$W = \frac{\mu}{r} \ , \ g = \frac{\mu}{r^2} \ ; \ r \geq R \quad (\mu = GM) \ , \tag{41}$$

and, by using Gauss's theorem,

$$W(r) = \frac{\mu}{R} + \int_r^R g(s)ds \ , \ g(r) = \frac{GM(r)}{r^2} \tag{42}$$

with

$$M(r) = 4\pi \int_0^r \rho(s)s^2 ds \ . \tag{43}$$

the total mass included in the sphere of radius r.

In both regions, in and outside the sphere of radius R, we have

$$dH = dr = -\frac{dW(r)}{g(r)} \ . \tag{44}$$

Therefore, it is only natural to pose the question whether this is the only case in which the level surfaces are parallel, the plumblines are straight and thus Eq. (18) holds true.

It is not difficult to see that the introduction of a centrifugal term into the gravity formula, however small the angular velocity, destroys the parallelism of the equipotential surfaces. As such, our question is only meaningful if we disregard rotation. We then answer the question whether it is possible that

$$dH = -\frac{dW}{g} \tag{45}$$

for a non-rotating potential, with the characteristics described in Conclusion 1, by applying the so-called *gravity space formalism* (cf. Sansò (1977)).

Therefore, we define an adjoint potential

$$\psi = \underline{x} \cdot \underline{g} - W \tag{46}$$

and we consider it as function of the gravity vector $\underline{g}$. If we differentiate Eq. (46), also known as the *Legendre transform*, we get

$$d\psi = d\underline{x} \cdot \underline{g} + \underline{x} \cdot d\underline{g} - dW = \underline{x} \cdot d\underline{g} \ , \tag{47}$$

proving that

$$\nabla_g \psi = \underline{x} \ . \tag{48}$$

Substituting Eq. (48) back into Eq. (46) we get

$$W = \underline{g} \cdot \nabla_g \psi - \psi = g \frac{\partial \psi}{\partial g} - \psi \ . \tag{49}$$

On the other hand, differentiating Eq. (48) once more, we obtain

$$\begin{aligned} \Psi = \left[\frac{\partial^2 \psi}{\partial g_i g_k} \right] = \frac{\partial \underline{x}}{\partial \underline{g}} = \left[\frac{\partial \underline{g}}{\partial \underline{x}} \right]^{-1} = \\ = W^{-1} = \left[\frac{\partial^2 W}{\partial x_i \partial x_k} \right]^{-1} \ . \end{aligned} \tag{50}$$

Accordingly, outside the masses, one has

$$Tr\Psi^{-1} = TrW = \Delta W = 0 \ . \tag{51}$$

On the other hand, if W has to be a function of g only, due to Eq. (49), we see that ψ also has to be function of g only. Then it is just a matter of an exercise to see that

$$\Psi = \frac{1}{g} \left[\psi' I - (\psi' - g\psi'') P_g \right] \tag{52}$$

with

$$P_g = \left[\frac{g_i g_k}{g^2} \right] \ , \ \psi' = \frac{\partial \psi}{\partial g} \ , \ \psi'' = \frac{\partial^2 \psi}{\partial g^2} \ . \tag{53}$$

It is then straightforward to verify that

$$\Psi^{-1} = \left[\frac{g}{\psi'} I + \left(\frac{1}{\psi''} - \frac{g}{\psi'} \right) P_g \right] \ . \tag{54}$$

Equation (51) then yields

$$Tr\Psi^{-1} = \frac{2g}{\psi'} + \frac{1}{\psi''} = 0 \tag{55}$$

or

$$2g\psi'' + \psi' = 0 \; . \tag{56}$$

To integrate Eq. (56), one has to take into account that when $r \to \infty$ the actual potential has the asymptotic behaviour

$$W \sim \frac{\mu}{r} \sim \mu^{1/2} g^{1/2} \; ; \tag{57}$$

then, from Eq. (52), one finds that when $g \to 0, \; \psi$ must have the asymptotic behaviour

$$\psi \sim -2\mu^{1/2} g^{1/2} \; . \tag{58}$$

The only integral of Eq. (56) having the behaviour of Eq. (58) at the origin is indeed

$$\psi = -2\mu^{1/2} g^{1/2} \tag{59}$$

itself. Therefore, the corresponding relation between $\underline{x}$ and $\underline{g}$ is

$$\underline{x} = -\mu^{1/2} g^{-3/2} \underline{g} \; ; \tag{60}$$

By inverting Eq. (60), one gets

$$\underline{g} = -\mu \frac{\underline{x}}{r^3} \; , \tag{61}$$

which proves that, at least outside the masses, the gravity field satisfying Eq. (45) has to be purely spherical.

As for the part inside the masses, we just cut short the reasoning; as a matter of fact if $W = W(g)$, inside as outside the masses, ψ also has to be function of g only too and Eq. (51) then becomes

$$Tr\Psi^{-1} = \frac{2g}{\psi'} + \frac{1}{\psi''} = -4\pi\rho(g) \; . \tag{62}$$

We do not know the exact analytical form of the integral of Eq. (62). However, the problem can be reduced to a quadrature for a layer of constant density, i.e.. to the integration of

$$\begin{aligned} \psi' = &- \left[\frac{g^3}{\rho^3} + \frac{3}{2}\,\frac{C}{\rho} + \sqrt{\frac{9C^2}{4\rho^2} + \frac{3Cg^3}{\rho^4}} \right] + \\ &-g^2 \left[\frac{g^3}{\rho} + \frac{3}{2} C\rho + \sqrt{\frac{9}{4}C^2\rho^2 + 3Cg^3} \right] - \frac{g}{\rho} \; , \end{aligned} \tag{63}$$

where C is a positive constant to be adapted to the boundary information that ψ' has to be equal to the radius of the bounding sphere. Whatever the integral of Eq. (64) is, we still must have, because of Eq. (48),

$$\underline{x} = \nabla\psi = \psi' \frac{\underline{g}}{g} \ . \tag{64}$$

Now, Eq. (64) implies

$$|\underline{x}| = r = \psi'(g) \Rightarrow g = F(r) \tag{65}$$

and

$$\underline{g} = F(r)\frac{\underline{x}}{r} \ , \tag{66}$$

which says again that the surfaces $g = \{\text{const}\}$, which coincide with level surfaces, have to be spheres.

Conclusion 4 The relation $dH = -\frac{dW}{g}$, which is not true for the actual Earth, cannot be satisfied by a rotating mass distribution, i.e., when g is the modulus of the gravity vector and not of gravitation only. Among purely gravitational fields, the above relation can only be satisfied by a field with simple spherical symmetry outside the masses (monopole field) and inside the masses down to the geoid; namely the "geoid" has to be a sphere and the corresponding density distribution has to be spherically layered.

5 Orthometric and other height systems, with some operative conclusions

Since other than orthometric height systems are used in geodesy, it seems logical to close this paper by mentioning them and by assessing their position with respect to the holonomity problem.

In particular, we can consider dynamic and normal heights (cf. Vaniček and Krakiwsky (1986)). As for dynamic heights the question is almost straightforward, because by definition

$$H^D \equiv \frac{W_0 - W_{(P)}}{\overline{\gamma}} \ , \tag{67}$$

where W_0 is a reference value (gravity potential on the geoid) and $\overline{\gamma}$ some reference constant value, close to the actual gravity modulus.

Indeed Eq. (67) implies

$$dH^D = -\frac{1}{\overline{\gamma}} dW \tag{68}$$

yielding

$$\int_\Gamma dH^D \equiv -\frac{1}{\overline{\gamma}} \int_\Gamma dW = 0 \tag{69}$$

for any closed loop Γ. Hence dynamic heights are holonomic.

As for normal heights, the problem is a little more intricate. Nevertheless, if we go back to the very definition, i.e., the normal height h^*_P is the ellipsoidal height of a point Q lying on the same normal to the ellipsoid as P and such that

$$U(h^*, \varphi, \lambda) = W(h, \varphi, \lambda) \ , \tag{70}$$

we see that, on account of the regularity of U as function of h^*, Eq. (70) can be inverted to provide

$$h^* \equiv h^*(h, \varphi, \lambda) \ . \tag{71}$$

Since h^* is smooth, due to the implicit function theorem (cf. Kaplan (1991)), we know that its differential will be exact on account of the argument used in §2. Therefore, the conclusion drawn for orthometric heights does not change too much with these systems. Of course for H^D and h^*, it is necessary to clarify how their differentials are related to the leveling increment δL, which is the observable quantity, in order to derive the expression of the corrections, as we have already done for the orthometric height in §3.

For H^D, the computation is particularly simple, since from Eq. (68) we can write (also see Eq. (24))

$$dH^D = -\frac{g}{\overline{\gamma}} \frac{dW}{g} = \frac{g}{\overline{\gamma}} \delta L = \delta L + \frac{g - \overline{\gamma}}{\overline{\gamma}} \delta L \ , \tag{72}$$

showing that the differential dynamic correction (DC) is given by

$$\delta H^{D*} = d(DC) \equiv \frac{g - \overline{\gamma}}{\overline{\gamma}} \delta L \ . \tag{73}$$

The integral form of Eq. (73) for a leveling line joining A to B is therefore

$$DC(AB) = \int_A^B \frac{g - \overline{\gamma}}{\overline{\gamma}} \delta L \ . \tag{74}$$

For the normal height, a similar, although slightly more complicated, expression can be derived (see Heiskanen and Moritz (1967), Ch. 4, §6).

Remark 6 It is interesting to establish a connection between $OC(AB)$ (Eq. 82) and $DC(AB)$ (Eq. 74), which is also pointed out in geodetic literature (e.g., Vaniček (1982)).

Let us introduce a path Γ defined as

$$\Gamma = (A_0'A) \cup (AB) \cup (BB_0') \ , \tag{75}$$

i.e., ascending along the ellipsoidal normal from A_0' to A, moving along the leveling line (AB) and then descending from B to B_0' along the new normal to the ellipsoid. We then recognize that

$$OC(AB) = \int_\Gamma \frac{\Delta g}{\gamma} \sin I d\Gamma \ , \tag{76}$$

where I is the inclination of the path element $d\Gamma$ on the horizontal plane;

$$\begin{aligned} &\text{on } (A_0'A) \ , \sin I = 1 \ , \quad d\Gamma = dh \\ &\text{on } (AB) \ , \ \ \sin I d\Gamma = \delta L \\ &\text{on } (BB_0') \ , \sin I = -1 \ , \quad d\Gamma = -dh \ . \end{aligned} \tag{77}$$

We now observe that a constant $\overline{\gamma}$ can always be chosen in such a way that

$$\int_\Gamma \frac{\Delta g}{\gamma} \sin I d\Gamma \cong \int_\Gamma \frac{\Delta g}{\overline{\gamma}} \sin I d\Gamma \equiv \int_\Gamma \frac{g - \overline{\gamma}}{\overline{\gamma}} \sin I d\Gamma \ . \tag{78}$$

Accordingly, the orthometric correction $OC(AB)$ can be decomposed as

$$OC(AB) = DC(A_0'A) + DC(AB) + DC(BB_0') \ , \tag{79}$$

which is the sought-after interpretation.

Remark 7 For any kind of height, we can always write equations like Eq. (22)

$$dH = \delta L + \delta H^* \ , \tag{80}$$

so that in a leveling network for any closed loop Γ, we can write a condition equation like

$$\int_\Gamma \delta L + \int_\Gamma \delta H^{D*} = 0 \ , \tag{81}$$

or similarly for the other heights.

In reality, however, we cannot use continuous equations, but rather have to write discrete sums along a closed leveling line with benchmarks $(P_1 \dots P_n)$, for which it is true that

$$\sum_{i=1}^{N} \delta H(P_i, P_{i+1}) = \sum_{i=1}^{N} \delta L(P_i, P_{i+1}) + \sum_{i=1}^{N} \delta H^*(P_i, P_{i+1})$$
$$= 0 \; , \tag{82}$$

with $P_{N+1} = P_1$.

Naturally, the integral of the correction δH^* along a side of the leveling network cannot be done exactly in general and only a discretized version of it can be used in Eq. (82), thus introducing a discretization error. Simultaneously, the use of observed leveling increments in Eq. (82) will also give a non-zero contribution due to measurement errors. In other terms, as we know, the use of real data will call for a least squares adjustment in order to reduce the influence of errors.

We come therefore to our last operative conclusion, taken from (Vaniček (1982)).

Conclusion 5 Any height system – dynamic, orthometric or normal – can be used in the adjustment of leveling networks with the same justification as the geopotential numbers. The only requirement is that the height differences needed for the adjustment be corrected for the effect of actual gravity observed on the surface of the Earth.

References

Andersen OB, Knudsen P (1982) Global marine gravity field from the ERS1 and GEOSAT geodetic mission altimetry. J Geophys Res 103(B4):8129-8137

Heiskanen WA, Moritz H (1967) Physical geodesy, Freeman, San Francisco, 364 pp

Grafarend EW (1975) Cartan frames and a foundation of physical geodesy. Methoden und Verfahren der Mathematischen Physic/Band 12, Brosowski B, Martensen E (eds), BI-Verlag, Mathematical Geodesy, Mannheim, pp 179-208

Grafarend EW, Suffys R, You RJ (1995) Prospective heigts in geometry and gravity space. Allgemeine Vermessungs-Nachrichten 102:382-403

Grafarend EW (1997) Field lines of gravity, their curvature and torsion, the Lagrange and the Hamilton equations of the plumbline. Annali di Geofisica 40(5):1233-1247

Hotine M (1969) Mathematical geodesy. ESSA Monograph 2, US Department of Commerce, Washington, 416 pp

Kaplan W (1991) Advanced calculus, 4th ed., Addison-Wesley, Reading, MA, 746 pp

MacMillan WD (1958) The theory of the potential, Dover Publications, New York

Marussi A (1985) Intrinsic geodesy, Springer-Verlag, Berlin, 240 pp

Miranda C (1970) Partial differential equations of elliptic type. second revised edition. Translated from Italian by Zane C. Motteler. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 2, Springer-Verlag, New York

Sansò F (1977) The geodetic boundary value problem in gravity space. Memorie Scienze Fisiche, Accademia Nazionale dei Lincei.

Vaníček P, Krakiwsky EJ (1986) Geodesy: the concepts, second edition, North-Holland Publishing, Amsterdam

Vaníček P (1982) To the problem of holonomity of height systems. Letter to the editor, The Canadian Surveyor 36:122-123

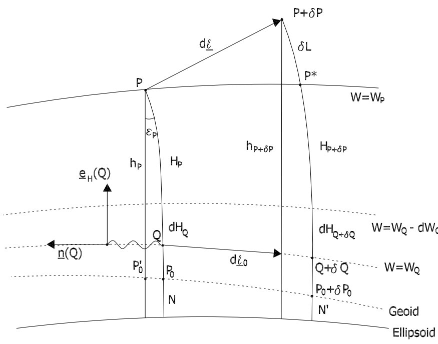

Fig. 1. Geometry of plumblines, equipotential surfaces, geoid (G) and ellipsoid (E); $h_P, h_{P+\delta P}$ ellipsoidal heights; δL leveling increment; $H_P, H_{P+\delta P}$ orthometric heights, i.e., arclenghts PP_0 and $P + \delta P P_0 + \delta P_0$; $dH_Q, dH_{Q+\delta Q}$ plumblines elements between two equipotential surfaces $W = W_Q, W = W_Q - dW_Q$; $W = W_P, W = W_Q, W = W_Q - dW$ equipotential surfaces; $d\underline{\ell}$ vector of the displacement $\delta P, d\underline{\ell}_0$ horizontal component of $d\underline{\ell}$; N, N' geoid undulations corresponding to $P_0, P_0 + \delta P_0$; ε_P deflection of the vertical and inclination of the section PP^* on the ellipsoid; $\underline{e}_H(Q), \underline{n}(Q)$ tangent and principal normal unit vectors of the plumbline H_Q.