Journal of Geodesy (1998) 72: 71–100

Journal of Geodesy

Generalized inverses of nonlinear mappings and the nonlinear geodetic datum problem

A. Dermanis

Department of Geodesy and Surveying, The Aristotle University of Thessaloniki, University Box 503, GR-54006 Thessaloniki, Greece
Phone: +30 31 996111; fax: +30 31 996137; email: dermanis@topo.auth.gr

Received: 11 April 1996 / Accepted: 19 April 1997

Abstract. Motivated by the existing theory of the geometric characteristics of linear generalized inverses of linear mappings, an attempt is made to establish a corresponding mathematical theory for nonlinear generalized inverses of nonlinear mappings in finite-dimensional spaces. The theory relies on the concept of fiberings consisting of disjoint manifolds (fibers) in which the domain and range spaces of the mappings are partitioned. Fiberings replace the quotient spaces generated by some characteristic subspaces in the linear case. In addition to the simple generalized inverse, the *minimum-distance* and the x_0-*nearest* generalized inverse are introduced and characterized, in analogy with the least-squares and the minimum-norm generalized inverses of the linear case. The theory is specialized to the geodetic mapping from network coordinates to observables and the nonlinear transformations (Baarda's S-transformations) between different solutions are defined with the help of transformation parameters obtained from the solution of nonlinear equations. In particular, the transformations from any solution to an x_0-nearest solution (corresponding to Meissl's inner solution) are given for two- and three-dimensional networks for both the similarity and the rigid transformation case. Finally the nonlinear theory is specialized to the linear case with the help of the singular-value decomposition and algebraic expressions with specific geometric meaning are given for all possible types of generalized inverses.

Key words. Generalised inverse · Nonlinearity · Pseudoinverse · Inverse problems · Datum problem

1 Introduction

The datum problem or zero-order design problem (Grafarend 1974) arising in the adjustment of observations related to geodetic networks has received considerable attention since the pioneering work of Meissl (1965, 1969) and its popularization by Blaha (1971). The nature of the problem has been clarified in two important papers by Grafarend and Schaffrin (1974, 1976) while the relation among various solutions, as well as Meissl's "inner" solution have been established with the introduction of the famous Baarda S-transformation (Baarda 1973; see also Mierlo 1980; Koch 1982). It might therefore seem that every aspect of this problem has been fully investigated and well understood for quite some time now, though some recent work (Xu 1995) might point to the contrary. However, with the exception of some simple cases, such as that of levelling in a small area, the mathematical treatment of the problem is confined to its linearized version, although it is well understood as a nonlinear problem. Nonlinear adjustment has been extensively studied in the geodetic literature, especially from the viewpoint of nonlinear least squares, see, e.g., Teunissen (1985, 1989a, 1989b, 1990), Grafarend and Schaffrin (1991) and Lohse (1994), where further references can be found. A completely different approach is that of Dermanis and Sanso (1995), where optimal nonlinear estimators have been investigated from a strictly probabilistic point of view. In both approaches the datum problem is solved in an implicit way. In the nonlinear least-squares adjustment solved by an iteration scheme, the choice of datum depends on the choice of the initial parameter values used for starting the iteration process, as well as on the principle used for "improving" these parameters. In the nonlinear estimation case, the need to adopt a Bayesian point of view in order to obtain a meaningful estimation independent of the unknown "true" parameter values also solves the datum problem in an implicit way, where the datum choice is hidden in the choice of the prior probability distribution of the parameters.

The solution to the nonlinear datum problem presented here is based on the concept of the S-transformation and has the form of such a similarity (or rigid) transformation with parameters which come from the solution of a system of nonlinear equations.

Another important aspect of the geodetic datum problem in its linear form is its relation to generalized inverses of matrices (linear operators) which led Bjerhammar (1951) to an independent introduction of the Moore-Penrose generalized inverse, later than Moore (1920) but before Penrose (1955). The question arises whether the nonlinear version of the geodetic datum problem bears a similar relation to some type of generalized inverses of nonlinear operators. An attempt will be made to look mainly into the geometric aspects of such nonlinear generalized inverses, although the building of a concrete mathematical theory requires a more rigorous treatment which is beyond our present scope. We shall base our investigation on the representation theory of various types of linear generalized inverses which has been introduced by Takos (1976) from an algebraic point of view and especially by Teunissen (1985) from a geometric point of view.

The datum problem is always a part of the problem of the adjustment of redundant observations which are related to a set of parameters (coordinates in the geodetic case) which in fact cannot be determined from the available observations. The reason is that the information contained in the angle and distance (angle) observations relates only to the shape and size (shape) of the network, while coordinates relate in addition to its position (position and size) with respect to a certain reference frame. The problem of placing the network in relation to a given reference frame can be also seen from the inverse point of view of placing a reference frame in relation to a given (i.e., physically existing) network. This choice of reference frame (or datum in geodetic terminology) poses a problem, the *datum problem*, which must be solved in an arbitrary but consistent way based in the introduction of additional information not contained in the observations.

The usual approach to the description of the adjustment and choice of datum problem is to consider the n observed parameters ($n > r$) as the coordinates of a linear space Y, called the *observation space*, in which the r-dimensional manifold M modeling the physical system, called the *model manifold*, is lying. (We confine ourselves to the case of discrete and finite observations.) The model manifold is described as the range of a nonlinear operator f from an m-dimensional *parameter space* X ($m > r$) into the observation space Y. The mapping f is established by the mathematical equations which relate all observables to the unknown parameters. In a "normal" situation, which is almost never the case in geodesy, the number of parameters and the dimension of the model manifold are equal ($m = r$), in which case the restriction $f_{|M}$ of f to M has an inverse which may serve as a coordinate mapping from M to $X = R^m$. In other words the chosen parameters can serve as a particular system of coordinates on the model manifold and the only problem to be solved is the adjustment problem. In the linear case the rank of f is $r = m$ and the corresponding model is called a *full-rank model*.

The unavoidable observational errors are added to the true values of the observables (which correspond to a point on the model manifold) and give observations corresponding to a known point outside the manifold. The *adjustment problem* can be simply defined as the problem of finding an optimal way to "return to the model manifold". In the case where $m > r$ (*model without full rank* in the linear case), the determination of parameter values is not trivial anymore, because there is an infinite set of parameter values which f maps on the same manifold point corresponding to the adjusted observations. The *datum problem* is exactly the problem of choosing one out of all possible parameter sets.

Of course rank deficiency in an observational model is not exclusively related to the datum problem, as demonstrated, e.g., in Dermanis and Grafarend (1981). Additional rank deficiency may result from the inability of the available observations to recover the shape (or shape and size) of a geodetic network. Such cases, although partly covered by the treatment in Sect. 3, need individual treatment and cannot be part of the general approach taken here, where our main subject is the common datum defect resulting only from the use of coordinates as parameters.

We must point out that the above point of view is not coordinate-free, since it depends on the choice of a specific set of m unknown parameters. This is perhaps of little concern to geodesy where the choice of coordinates as parameters imposes itself as a matter of convenience. It is possible, however, to establish a theory where both the adjustment and the datum problem are treated in a coordinate-free way in the spirit of modern differential geometry.

A more in-depth introduction to the (nonlinear) datum problem, especially in relation to the modelling problem, can be found in Dermanis (1991).

2 Geometric characteristics of generalized inverses of linear operators

A natural point of departure for the study of the nonlinear datum problem is our knowledge of the simpler linear case. For this reason we shall review some of the geometric characteristics of the theory of generalized inverses of linear mappings and point out those which have proven to be more appropriate for generalization to the nonlinear case. Our exposition will be more casual than in the rest of this work, since a rigorous and thorough study exists already in the geodetic literature (Teunissen 1985).

A linear mapping $f : X \to Y$ is characterized geometrically by its range

$$R(f) = \{y \in Y | y = f(x) \text{ for some } x \in X\}$$

which is a linear subspace of Y and its null space

$$N(f) = \{x \in X | f(x) = 0\}$$

which is a linear subspace of X. To any element $y = f(x)$ of $R(f)$ corresponds an affine subspace $x + N(f)$ of X with elements which f maps to the same element $y = f(x)$. The parallel translates of $N(f)$ are thus the "solution spaces" of f which correspond one-to-one to the elements of $R(f)$.

If X is m-dimensional, Y is n-dimensional and rank $(f) = \dim R(f) = r$ then $d = m - r$ is the injectivity defect of f, while $f = n - r$ is its surjectivity defect (usually called degrees of freedom). Furthermore $N(f)$ is a d-dimensional subspace of X and $R(f)$ is an r-dimensional subspace of Y.

Coming to a linear generalized inverse g of f, it follows from the defining property $f = f \circ g \circ f$ that both the linear operators $p = f \circ g$ and $q = g \circ f$, are idempotent and thus linear projectors. As such they are completely characterized by their range (invariant linear subspace) and their complementary null-space (linear subspace along which they project), see e.g. Halmos (1974, Sect. 41), Schaffrin et al. (1977).

We shall first show that $R(p) = R(f)$. For every $y \in Y$, $\hat{y} \equiv p(y) = f(g(y)) \in R(f)$, implying that $R(p) \subset R(f)$. On the other hand, for any $\hat{y} \in R(f)$ there exists $x \in X$ such that $\hat{y} = f(x)$ and $p(\hat{y}) = (f \circ g)(f(x)) = (f \circ g \circ f)(x) = f(x) = \hat{y}$, implying that $R(f) \subset R(p)$. Therefore $R(p) = R(f)$.

$R(p) = R(f)$ is an r-dimensional (linear) subspace of Y. The linear projector p is completely determined if also the f-dimensional subspace $C \subset Y$ is given such that $Y = R(f) \oplus C$ and $y - p(y) \in C$ for every $y \in Y$. Thus p is a linear projection on $R(f)$ along C.

The elements of C are projected by p to the zero element of Y, i.e., $C = N(p) = N(f \circ g)$.

Indeed for any $y \in C$ it must hold that $\hat{y} = p(y) \in R(f)$ while $y - \hat{y} \in C$. But C is a linear subspace, and if $y \in C$ and $y - \hat{y} \in C$ so is their difference, i.e., $y - (y - \hat{y}) = \hat{y} \in C$.

Now $\hat{y} \in C$ and $\hat{y} \in R(f)$, while $C \cap R(f) = \{0\}$, and therefore $\hat{y} = 0$.

From the fundamental property $f = f \circ g \circ f$ and the well-known property for the rank of matrix product (mapping composition) it follows that $r_g = \text{rank}(g)$ cannot be less than the rank r of f, i.e., $r_g \geq r$. $N(g)$ is thus a subspace of Y with dimension $f_g = n - r_g \leq n - r = f$.

We shall show that in fact $N(g) \subset C$: if $y \in N(g)$ then $g(y) = 0$ and $p(y) = f(g(y)) = f(0) = 0$, i.e., also $y \in N(p) = C$ and therefore $N(g) \subset C$.

For every $\hat{y} \in R(f)$ the affine subspace $\hat{y} + C$ consists of all elements of Y which p maps on the same element $\hat{y}$. The elements of any particular affine subspace $y + N(g)$ are mapped on the same element $x = g(y) \in R(g) \subset X$. Further mapping of this particular element x by f to $\hat{y} = f(x)$ has as a consequence that all elements of $y + N(g)$ are mapped by $p = f \circ g$ on the same element $\hat{y} = p(y)$. Therefore $y + N(g) \subset p(y) + C = y + C$, and furthermore $N(g) \subset C$ as already shown. A similar situation holds for the linear projector q. For any $x \in X$, $\hat{x} = q(x) = (g \circ f)(x) \in g(R(f)) \subset R(g)$.

The subspace $S = R(q) = g(R(f)) \subset R(g)$ is the subspace onto which q maps elements of X. If two elements x_1, x_2 of X are mapped onto the same element $\hat{x} \in S$, then $\hat{x} = q(x_1) = q(x_2)$ implies that $g(f(x_1) - f(x_2)) = 0$ and $f(x_1) - f(x_2) \in N(g) \subset C$. But due to $Y = R(f) \oplus C$ we have $R(f) \cap C = \{0\}$ and $R(f) \cap N(g) = \{0\}$. Since $f(x_1) - f(x_2) = f(x_1 - x_2)$ belongs to both $R(f)$ and $N(g)$ it must hold that $f(x_1) - f(x_2) = 0$ and $x_1 - x_2 \in N(f)$.

Conversely, if $x_1 - x_2 \in N(f)$ then $f(x_1) - f(x_2) = f(x_1 - x_2) = 0$, implying $q(x_1) - q(x_2) = g(f(x_1 - x_2)) =$

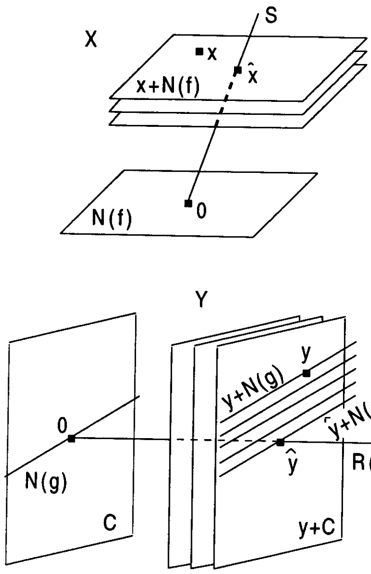

Fig. 1. The geometry of the generalized inverse g of a linear mapping f. Here $x = g(y)$, $\hat{y} = f(x) = p(y)$, while $\hat{x} = g(\hat{y}) = q(x)$. All the elements of $\hat{y} + N(g)$ are mapped into x, while all the elements of $y + C$ are mapped into $\hat{x}$

$g(0) = 0$ and all elements of any affine subspace $x + N(f)$ are projected by q on the same element $\hat{x}$.

Therefore q projects elements of X on $S = g(R(f))$ along $N(f)$ and $X = N(f) \oplus S$. Since $N(f)$ is d-dimensional, S will be an r-dimensional subspace of X.

An important property of S is that it intersects any solution space (affine subspace) $x + N(f)$ of f in a single element.

To see this let both $\hat{x}_1$ and $\hat{x}_2$ belong to $S \cap [x + N(f)]$. Since $\hat{x}_1 \in [x + N(f)]$, $x - \hat{x}_1 \in N(f)$, $f(x - \hat{x}_1) = 0$ and $q(x - \hat{x}_1) = g(f(x - \hat{x}_1)) = g(0) = 0$, implying that $q(\hat{x}_1) = q(x)$. Since $\hat{x}_1 \in S$ it holds that $q(\hat{x}_1) = \hat{x}_1$ and thus $\hat{x}_1 = q(x)$. With similar reasoning $\hat{x}_2 = q(x)$ and therefore $\hat{x}_1 = \hat{x}_2$.

$R(f)$ and S are subspaces of the same dimension which are in a one-to-one correspondence. If $\hat{y} \in R(f)$ then $\hat{y} = f(x)$ for some $x \in X$ and $g(\hat{y}) = g(f(x)) = q(x) \equiv \hat{x} \in S$. Also $\hat{x} \in [x + N(f)]$ and $\hat{x} = g(\hat{y}) = q(x)$ is the unique element in the intersection of $x + N(f)$ and S, i.e., $\{\hat{x}\} = [x + N(f)] \cap S$. For any element $\hat{x} \in S$ there corresponds a unique element $\hat{y} \in R(f)$ such that $\hat{y} = f(\hat{x})$ and $\hat{x} = g(\hat{y})$. We may say that the restriction of g to $R(f)$ is the ordinary inverse of the restriction of f on S: $g_{|_{R(f)}} = (f_{|_S})^{-1}$. In symmetry with $S = g(R(f))$ it holds also that $R(f) = f(S)$.

Once S is given g is defined on $R(f)$. It remains to be defined outside $R(f)$. Let $y \in R(f)$ and $\hat{y} = p(y)$. Then $y - \hat{y} \in C$. Taking into account that $N(g) \subset C$ we can decompose

$$y - \hat{y} = (y - \hat{y})_N + (y - \hat{y})_{C'}$$

into a part $(y - \hat{y})_N \in N(g)$ and a part $(y - \hat{y})_{C'} \in C'$ where C' is a complement of $N(g)$ with respect to C: $C = N(g) \oplus C'$. This decomposition is not unique but depends on the specific choice of the subspace $C' \subset C$

which is of dimension $\dim C' = \dim C - \dim N(g) = f - (n - r_g) = n - r - (n - r_g) = r_g - r \equiv \Delta r$.

On the basis of the decomposition

$$y = \hat{y} + (y - \hat{y}) = \hat{y} + (y - \hat{y})_N + (y - \hat{y})_{C'}$$

and the linearity of g we have:

$$g(y) = g(\hat{y}) + g((y - \hat{y})_N) + g((y - \hat{y})_{C'}) = g(\hat{y}) + g((y - \hat{y})_{C'}).$$

The $g(\hat{y})$ part is uniquely defined: the subspaces $R(f)$ and C determine $\hat{y}$ and $g(\hat{y})$ is the unique element in the intersection of S with the affine subspace $x + N(f)$ which f maps into $\hat{y}$ (the solution space of $\hat{y}$). It remains to determine how g acts on the subspace C'. This is the only required information to define completely the action of g in addition to a description of S and C. If a basis is known for $N(g)$ and a basis for C' they constitute together a basis of C which is thus defined. The generalized inverse g is specified if (in addition to a basis for S) we know how it acts on all the $r_g - r$ elements of the basis of C'.

A unique choice of C' can be specified without ambiguity (if so desired) by requiring in addition that C' is the orthogonal complement of $N(G)$ with respect to C, in which case C assumes an orthogonal decomposition $C = N(g) \stackrel{\perp}{\oplus} C'$.

Let $D = g(C)$ be the image of C under g. If $y \in C$ we have the unique decomposition $y = y_N + y_{C'}$ where $y_N \in N(G)$ and $y_{C'} \in C'$. In this case $g(y) = g(y_N) + g(y_{C'}) = g(y_{C'})$ and $g(C) = g(C') = D$. If $y \in C = N(p)$ then $p(y) = 0 = (f \circ g)(y) = f(g(y))$ and $g(y) \in N(f)$ so that $g(C) = D \subset N(f)$.

The generalized inverse g is a one-to-one mapping from C' to D and $\dim D = \dim C' = \Delta r \le d = \dim N(f)$, the last inequality following from the fact that $D \subset N(f)$.

The range $R(g) = g(Y)$ can be decomposed according to $Y = R(f) \oplus C$ as $R(g) = g(R(f)) \oplus g(C)$, and since $g(R(f)) = S$ and $g(C) = D$, we have that $R(g) = S \oplus D$.

We introduce now an alternative perspective to the geometry of a generalized inverse which might seem at first to be unnecessarily complicated, but it will prove helpful in the study of the nonlinear case.

To define g on Y it is sufficient to define it on a subspace M complementary to $N(g)$, since for $Y = N(g) \oplus M$ we have a unique decomposition $y = y_N + y_M$ with $y_N \in N(g)$, $y_M \in M$ and $g(y) = g(y_M)$. Such a subspace M is provided by $M = C' \oplus R(f)$. If $y \in M$ with unique decomposition $y = y_{C'} + y_R$, $y_{C'} \in C', y_R \in R(f)$ then $g(y) = g(y_{C'}) + g(y_R)$, where $g(y_{C'}) \in D$ and $g(y_R) \in S$. The affine subspace $y + C' = y_R + C'$ of M is mapped into $g(y_R) + D = g(y) + D$ which is an affine subspace of $R(g)$. The affine subspace $y + R(f)$ of M is mapped into an affine subspace $g(y) + S = g(y_{C'}) + S$ of $R(g)$. The image $g(y)$ is the unique intersection of these two affine subspaces of $R(g)$. Since $g(y) + D = g(y_R) + D$ is uniquely determined from the y_R component of y, it remains to specify a correspondence Γ from the affine subspaces $y + R(f)$ of M to the affine subspaces $g(y) + S$ of $R(f)$. If $g(y) + S = \Gamma(y + R(f))$ then $g(y)$ is determined from the intersection

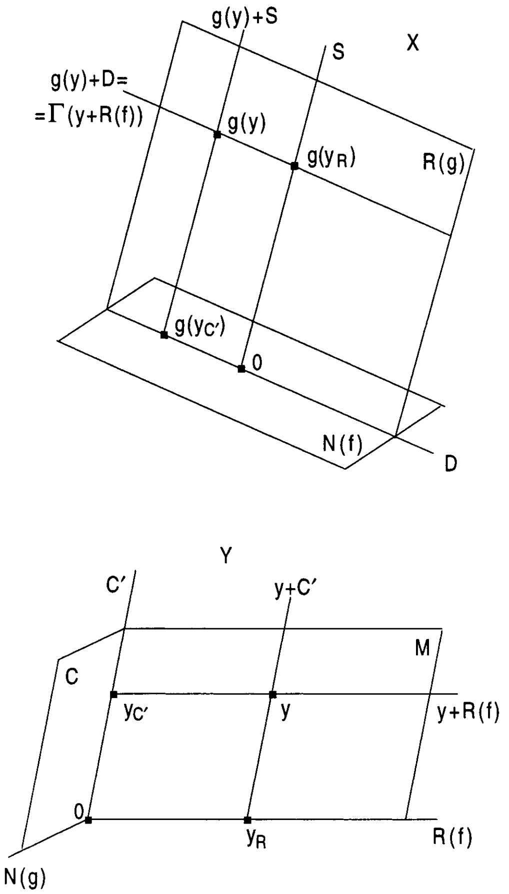

Fig. 2. More geometry of the generalized inverse of a linear mapping: the intersection of $y + R(f)$ and $y + C'$ determines the unique point y which g (the generalized inverse of f) maps into $x = g(y)$. The image $g(y)$ is uniquely determined from the intersection of $g(y) + S$ (= image of $y + C'$) and $g(y) + D$ (= image of $y + R(f)$ under Γ)

$$[g(y_R) + D] \cap \Gamma(y + R(f)) = \{g(y)\}$$

In order to specify a unique g out of the class of all possible generalized inverses of f, we must specify the subspaces C complementary to $R(f)$, S complementary to $N(f)$ and $D \subset N(f)$. Furthermore we must specify the restricted linear mapping $g|_{C'}: C' \to g(C') = D$ which brings the elements of C' into a one-to-one correspondence with those of D.

Let us look next into the special case where g is a reflexive generalized inverse, i.e., when $g \circ f \circ g = g$ also. In this particular case $r_g = r$ necessarily ($r_g \ge r$ follows from the fundamental property and $r_g \le r$ from the reflexivity property). It follows that $S = R(g)$ and $C = N(g)$. Thus $p = f \circ g$ is a projector on $R(f)$ along $N(g)$ and $q = g \circ f$ is a projector on $S = R(g)$ along $N(f)$. Let us notice that a reflexive generalized inverse is completely determined once the subspaces C and S are known.

A least-squares generalized inverse is one with $C = R(f)^{\perp}$, in which case we have the orthogonal decomposition $Y = R(f) \stackrel{\perp}{\oplus} R(f)$, and $\hat{y} = p(y)$ is the closest element to y from $R(f)$. The projector p is the orthogonal projector on $R(f)$.

A minimum-norm generalized inverse is one for which $S = N(f)^{\perp}$, in which case $g(y)$ is the element of the solution affine subspace corresponding to $\hat{y} = p(y)$ which has the smallest norm. The projector q is the orthogonal projector on $N(f)^{\perp}$.

From all the preceding geometric characteristics of a generalized inverse $g: Y \to X$ of a linear mapping $f: X \to Y$, we summarize those which will be helpful in dealing with nonlinear mappings.

The space Y is "sliced" by $p = f \circ g$ into affine subspaces (p-slices) which are parallel to a specific (linear) subspace C. Each slice corresponds to a particular element $\hat{y}$ of $R(f) \subset Y$, since $p = f \circ g$ maps all the elements of the slice onto that $\hat{y}$.

The mapping f "slices" the space X into affine subspaces (f-slices) which are parallel to the subspace $N(f)$. Each slice corresponds to a particular element $\hat{y}$ of $R(f)$ since f maps all the elements of the slice on that $\hat{y}$, we may therefore call this slice "solution space of $\hat{y}$".

The generalized inverse g maps $R(f)$ to a subset S of $R(g)$ which has the property that it intersects each solution space (f-slice) in a single element.

The mapping $q = g \circ f$ slices X into q-slices, each slice corresponding to a particular element $\hat{x}$ of S, since q maps all the elements of the slice on that $\hat{x}$. The q-slices are identical to the f-slices.

Finally g "slices" Y into affine subspaces which are parallel to the subspace $N(g)$. These g-slices have the property that each is contained in one of the p-slices. Thus each p-slice is itself further sliced in a more "refined" set of g-slices. Only in the case of a reflexive generalized inverse are the p-slices and g-slices identical.

These slices, all of which are affine subspaces, can be described by a single "generating" linear subspace (f-slices by $N(f)$, g-slices, and q-slices by $N(g)$, p-slices by C). Of course, this possibility cannot survive in the nonlinear case.

3 Generalized inverses of nonlinear operators

Let X, Y be finite-dimensional spaces and $f: X \to Y$ a (nonlinear) mapping from X to Y. The question arises whether it is possible to define, under some appropriate conditions, a class of mappings $g: Y \to X$ which can serve as generalized inverses of the mapping f. The similar theory developed for linear mappings may serve as a guide to a certain extend, although the nonlinearity of f poses problems which do not allow for a direct extension of the existing theory.

Definition. A partition of a set A into disjoint sets A_t, $A = \cup_t A_t$ it is called a *fibering* of A, and its elements A_t are called *fibers*. Such a partition can arise from an equivalence relation on A, each fiber consisting of equivalent elements of A (Loomis and Sternberg 1980.)

A mapping f defined on X gives rise to a fibering $\mathscr{F}$ of X through the equivalence relation $x_1 \sim x_2$ if $f(x_1) = f(x_2)$. To every element $y \in R(f)$ in the range $R(f)$ of f corresponds a fiber $F_y \in \mathscr{F}$ defined by

$$F_y = \{x \in X | f(x) = y\} \tag{1}$$

The function f gives rise to a bijection

$$\bar{f}: R(f) \to \mathscr{F} : y \to F_y$$

Obviously for every $x \in X,\ x \in F_{f(x)}$.

Definition. The mapping $\pi: X \to \mathscr{F} : x \to F_{f(x)}$ is called the *projection mapping* from X to the fibering $\mathscr{F}$. It holds that $\pi = \bar{f} \circ f$.

When more than one fiberings are involved we denote π by $\pi_{\mathscr{F}}$. Thus $\pi_{\mathscr{F}}(x)$ is the unique fiber from the fibering $\mathscr{F}$ which passes through a given point x.

Definition. A *section* of a fibering $\mathscr{F}$ of X is the range $S = R(s)$ of a mapping $s: \mathscr{F} \to X$ such that $\pi \circ s = \mathrm{id}_{\mathscr{F}}$, where $\mathrm{id}_{\mathscr{F}}$ is the identity mapping in $\mathscr{F}$.

A section S of a fibering $\mathscr{F}$ intersects any one of its fibers in a single element: for any $x \in S, S \cap \pi_{\mathscr{F}}(x) = \{x\}$.

Definition. Let $\mathscr{F}$ and $\mathscr{K}$ be two fiberings of the same set A. $\mathscr{K}$ is called a *refinement* of $\mathscr{F}$ if every fiber K_t of $\mathscr{K}$ is a subset of some fiber F_s of $\mathscr{F} : \forall K_t \in \mathscr{K} : \exists F_s \in \mathscr{F}$ such that $K_t \subset F_s$.

If $\mathscr{K}$ is a refinement of $\mathscr{F}$ then $\pi_{\mathscr{K}}(z) \subset \pi_{\mathscr{F}}(z)$ for any z.

Definition. Two fiberings $\mathscr{F}$ and $\mathscr{H}$ of the same space X are called *complementary* if every fiber of $\mathscr{F}$ is a section of $\mathscr{H}$, and vice versa: $\pi_{\mathscr{F}}(x) \cap \pi_{\mathscr{H}}(x) = \{x\},\ \forall x \in X$.

Definition. A fibering $\mathscr{F}$ of X induces a fibering $\mathscr{F}' = \mathscr{F}_{|M}$ on a manifold $M \subset X$ with fibers $\pi_{\mathscr{F}'}(x) = M \cap \pi_{\mathscr{F}}(x)$ for every $x \in M$. The fibering $\mathscr{F}' = \mathscr{F}_{|M}$ is called the *restriction* of the fibering $\mathscr{F}$ on M.

Example 1. We shall use the simplest possible geodetic network in order to illustrate some of the results. Following Grafarend and Schaffrin (1974) we consider a three-point horizontal network $P_1P_2P_3$ with coordinates

$$x = [x_1\ y_1\ x_2\ y_2\ x_3\ y_3]^{\mathrm{T}}$$

where all sides $a = P_2P_3,\ b = P_1P_3,\ c = P_1P_2$, and all angles A, B, C at points P_1, P_2, P_3, respectively, have been observed. The observation vector

$$y = [a\ b\ c\ A\ B\ C]^{\mathrm{T}}$$

is a function $y = f(x)$ of the coordinates, explicitly given by

$$a = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$$

$$b = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}$$

$$c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$A = \arctan\frac{(x_3 - x_1)}{(y_3 - y_1)} - \arctan\frac{(x_2 - x_1)}{(y_2 - y_1)}$$
$$B = \arctan\frac{(x_2 - x_1)}{(y_2 - y_1)} - \arctan\frac{(x_3 - x_2)}{(y_3 - y_2)}$$
$$C = \arctan\frac{(x_3 - x_2)}{(y_3 - y_2)} - \arctan\frac{(x_3 - x_1)}{(y_3 - y_1)}$$

or

$$A = \arccos\frac{(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$
$$B = \arccos\frac{(x_3 - x_2)(x_1 - x_2) + (y_3 - y_2)(y_1 - y_2)}{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}$$
$$C = \arccos\frac{(x_3 - x_2)(x_3 - x_1) + (y_3 - y_2)(y_3 - y_1)}{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}}$$

We now consider the range $R(f)$. When $y = f(x)$ for some x, i.e., when $y \in R(f)$, then the six elements of y are not independent. Only three parameters, say $a,\ b,\ c$, are independent. The remaining ones, the angles, are uniquely determined from well-known trigonometric relations, e.g.,

$$\cos A = \frac{b^2 + c^2 - a^2}{2\ b\ c},$$
$$\cos B = \frac{a^2 + c^2 - b^2}{2\ a\ c},$$
$$\cos C = \frac{a^2 + b^2 - c^2}{2\ a\ b}$$

This means that $R(f)$ is of dimension $6 - 3 = 3$, since the three cosine conditions satisfied by its elements reduce the number of independent parameters by three. We may think of $q = [abc]^{\mathrm{T}}$ as three curvilinear coordinates on $R(f)$ which in this case obtains the direct representation $y = y(q)$, explicitly given by

$$a = a, \quad b = b, \quad c = c$$

$$A = \arccos\frac{b^2 + c^2 - a^2}{2\ b\ c},$$
$$B = \arccos\frac{a^2 + c^2 - b^2}{2\ a\ c},$$
$$C = \arccos\frac{a^2 + b^2 - c^2}{2\ a\ b}$$

We now move on to the solution space F_y for a fixed $y \in R(f),\ x \in F_y$ whenever $f(x) = y$. Since A, B, C are determined from a, b, c, the only independent conditions on x are those related to the given values of a, b, c :

$$a = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$$
$$b = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}$$
$$c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

These three conditions reduce the dimension of F_y to $6 - 3 = 3$. If x_0 is a fixed element of X the solution space through x_0 is the fiber $\pi_{\mathscr{F}}(x_0)$ and $x \in \pi_{\mathscr{F}}(x_0)$ if it satisfies the three conditions

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = (x_{03} - x_{02})^2 + (y_{03} - y_{02})^2$$
$$(x_3 - x_1)^2 + (y_3 - y_1)^2 = (x_{03} - x_{01})^2 + (y_{03} - y_{01})^2$$
$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = (x_{02} - x_{01})^2 + (y_{02} - y_{01})^2$$

which correspond to $f(x) = f(x_0)$ expressed by $a = a_0$, $b = b_0, c = c_0$. An alternative equivalent set of conditions is

$$bc \cos A = b_0 c_0 \cos A_0$$
$$ac \cos B = a_0 c_0 \cos B_0$$
$$ab \cos C = a_0 b_0 \cos C_0$$

which in terms of coordinates are

$$(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)$$
$$\quad = (x_{03} - x_{01})(x_{02} - x_{01}) + (y_{03} - y_{01})(y_{02} - y_{01})$$
$$(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)$$
$$\quad = (x_{02} - x_{01})(x_{03} - x_{01}) + (y_{02} - y_{01})(y_{03} - y_{01})$$
$$(x_3 - x_2)(x_1 - x_2) + (y_3 - y_2)(y_1 - y_2)$$
$$\quad = (x_{03} - x_{02})(x_{01} - x_{02}) + (y_{03} - y_{02})(y_{01} - y_{02})$$

We now return to the theory. Instead of speaking of the fibering $\mathscr{F}$ of X induced by a certain mapping f, we may take the inverse approach and describe a given fibering by means of a mapping. For a mapping to induce a fibering on its domain space it is necessary that it is not injective, otherwise we obtain the "trivial fibering" where each fiber consists of a single element of X!

The use of a mapping which, as our mapping of interest f, is not surjective, is not the most efficient way for describing a fibering. More straightforward is the use of a surjective mapping, say $h: X \to R^q$, where to every element $d \in R^q$ corresponds a fiber $H_d = \{x \in X | h(x) = d\}$ from the fibering $\mathscr{H}$ induced by h. The fibers H_d are manifolds of dimension $m - q = \dim X - q$. A more general way of describing a fibering is by means of a mapping

$$\chi: X \times R^q \to R^q: (x, d) \to \chi(x, d)$$

with $q < m$, such that $\chi_d(\cdot) \equiv \chi(\cdot, d)$ is surjective for every $d \in R^q$. This is guaranteed when $\left|\frac{\partial \chi}{\partial d}\right| \neq 0$ at every x and d. The fibers are the $(m - q)$-dimensional manifolds

$$H_d = \{x \in X | \chi(x, d) = 0\}$$

The previous case corresponds to the choice $\chi(x, d) = h(x) - d$.

Another approach to the description of fiberings of a space X is by means of mappings having X as their range rather than their domain as before. A bijective mapping

$$\phi: R^q \times R^{m-q} \to X: (d, p) \to \phi(d, p)$$

describes a fibering $\mathscr{H}$ with fibers $H_d = \{x \in X | x = \phi(d, x)\}$. Each fiber is the range $H_d = R(\phi_d)$ of the mapping $\phi_d(\cdot) = \phi(d, \cdot)$. The inverse mapping ϕ^{-1} is in fact a coordinate mapping on X where the coordinates $(d,p) = \phi^{-1}(x)$ are "adapted" to the fibering: each fiber corresponds to a fixed value of d, while the "free" parameters p may serve as (intrinsic) coordinates on the fiber.

We assume for the moment that f is defined on the whole of X, it is continuous, and has a differential mapping $\mathrm{d}f_x$ at each x which is a linear mapping from the tangent vector space T_x at x to the tangent vector space $T_{f(x)}$ at $f(x)$ (Choquet-Bruhat et al. 1977, p. 121). The mapping df_x is defined by

$$\begin{aligned} &\mathrm{d}f_x : T_x \to T_{f(x)} : v \to u \\ &u(f \circ \phi) = v(\phi) \quad \text{for every } f : X \to R \end{aligned} \tag{2}$$

(where the tangent vectors v and u are visualized as "directional derivatives" acting on functions).

We further assume that df_x has constant rank over X, $\mathrm{rank}(df_x) = r \leq \min(n,m)$, where $m = \dim X$ and $n = \dim Y$. Consequently $R(f)$ is an r-dimensional submanifold of Y.

We now come to the possibility of defining a generalized inverse g of the mapping f by a similar way as in the linear case.

Definition. A mapping $g: Y \to X$ is called a *generalized inverse* of a given mapping $f: X \to Y$ when

$$f \circ g \circ f = f \quad (3) \tag{G1}$$

[We denote Eq. (3) by (G1).] This means that for every $x \in X, (f \circ g \circ f)(x) = f(x)$, so that for every $y = f(x) \in R(f)$ it holds that $(f \circ g)(y) = f(g(y)) = y$. As a consequence $g(y) \in F_y$, i.e., g must map every element of $R(f)$ into an element of its corresponding fiber. In other words g maps $R(f)$ onto a section S of the fibering $\mathscr{F}$ where $S \subset R(g)$. The restriction of g on $R(f)$ is a bijection between $R(f)$ and S which are both r-dimensional manifolds.

The relation (G1) satisfied by the generalized inverse g implies a corresponding relation between the differentials $\mathrm{d}f_x$ and $\mathrm{d}g_{f(x)}$ of f and g respectively:

$$\mathrm{d}f_x \circ \mathrm{d}g_{f(x)} \circ \mathrm{d}f_x = \mathrm{d}f_x \tag{4}$$

which follows by implementing the implicit function theorem (Choquet-Bruhat et al. 1977, p. 91).

Lemma. $R(f \circ g) = R(f)$.

Proof. $y \in R(f) \Rightarrow \exists\, x \in X$:

$$\begin{aligned} y &= f(x) = (f \circ g \circ f)(x) = (f \circ g)[f(x)] \\ &= (f \circ g)(y) \in R(f \circ g) \Rightarrow R(f) \subset R(f \circ g) \end{aligned}$$

which combined with the obvious relation $R(f \circ g) \subset R(f)$ implies that in fact $R(f \circ g) = R(f)$. □

Assuming that g has a constant rank $r_g = \mathrm{rank}(dg_y)$ for every y, application of the property of the rank of a linear map composition $r(AB) \leq \min[r(A), r(B)]$ to Eq. (3), gives $r \leq \min(r, r_g) \leq r_g$ so that

$$r \leq r_g \leq \min(n,m) \tag{5}$$

Once S is given, g is defined on $R(f) \subset Y$. The question is how this particular g is extended outside $R(f)$. A step in this direction is to note that as a consequence of (G1) both $q = g \circ f$ and $p = f \circ g$ are idempotent mappings $(q^2 = q, p^2 = p)$

$$\begin{aligned} f \circ g \circ f = f \Rightarrow (g \circ f) \circ (g \circ f) = g \circ f \text{ and} \\ (f \circ g) \circ (f \circ g) = f \circ g \end{aligned} \tag{6}$$

which can be considered as "nonlinear projections"

$$\begin{aligned} &q = g \circ f : X \to R(g \circ f) \subset R(g) \subset X \\ &p = f \circ g : Y \to R(f \circ g) = R(f) \subset Y \end{aligned} \tag{7}$$

We have set $R(g \circ f) \subset R(g)$, which is obvious, and $R(f \circ g) = R(f)$ according to the previous lemma.

Elements belonging to the range of idempotent mappings are invariant under the mapping. Idempotent mappings induce also a fibering of the space they act on. We denote by $\mathscr{P}$, $\mathscr{Q}$ the fiberings induced by p and q, respectively, with corresponding elements $P_{y,y} \in R(p) = R(f)$ and $Q_{x,x} \in R(q) \subset R(g)$.

The mapping g gives rise to a fibering $\mathscr{G}$ of Y with elements $G_x = \{y \in Y | g(y) = x\}$. Obviously for every element y of Y it holds that $y \in G_{g(y)} = \pi_{\mathscr{G}}(y)$. If $y \in R(f)$ then $y \in G_x = \pi_{\mathscr{G}}(f(x))$, where $\{x\} = F_y \cap S$.

Lemma. For the fiberings of the space Y it holds that the fibering $\mathscr{G}$ induced by g is a refinement of the fibering $\mathscr{P}$ induced by $p = f \circ g$.

Proof. For any $z \in Y, \pi_{\mathscr{G}}(z) \in \mathscr{G}$ and $\pi_{\mathscr{P}}(z) \in \mathscr{P}$. For every $y \in \pi_{\mathscr{G}}(z), g(y) = g(z) \Rightarrow p(y) = f(g(y)) = f(g(z)) = p(z)$ and $y \in \pi_{\mathscr{P}}(z)$. Consequently $\pi_{\mathscr{G}}(z) \subset \pi_{\mathscr{P}}(z)$ for any $z \in Y$ and thus $\mathscr{G}$ is a refinement of $\mathscr{P}$. □

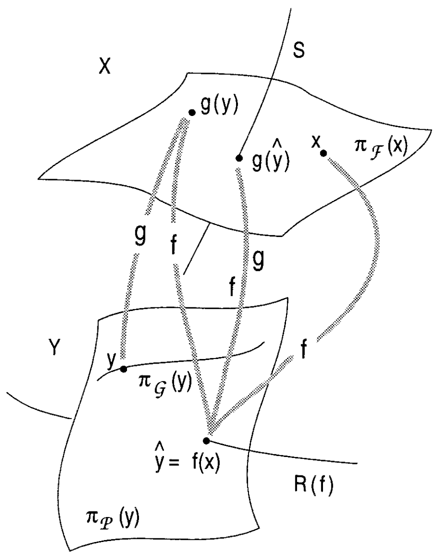

Fig. 3. The geometry of the generalized inverse g of a nonlinear mapping f. All elements of $\pi_{\mathscr{G}}(y)$ are mapped into the same element $g(y)$. The elements of $\pi_{\mathscr{P}}(y)$ are mapped into elements of $\pi_{\mathscr{F}}(x)$ and thus projected by p on the same $\hat{y}$

Example 1 (continued). We claim that the mapping g defined by $x = g(y)$, explicitly

$$x_1 = \frac{c - b\sin A}{3} \qquad y_1 = \frac{a - c - b\cos A}{3}$$

$$x_2 = \frac{c - b\sin A}{3} \qquad y_2 = \frac{a + 2c - b\cos A}{3}$$

$$x_3 = \frac{c + 2b\sin A}{3} \qquad y_3 = \frac{a - c + 2b\cos A}{3}$$

is a generalized inverse of the previously defined mapping f. To prove this we need to show that, for any x, $y = f(x)$ is identical to $y' = (f \circ g \circ f)(x) = (f \circ g)(y) = f(x')$, where $x' = g(y)$ is explicitly given by

$$x'_1 = \frac{c - b\sin A}{3} \qquad y'_1 = \frac{a - c - b\cos A}{3}$$

$$x'_2 = \frac{c - b\sin A}{3} \qquad y'_2 = \frac{a + 2c - b\cos A}{3}$$

$$x'_3 = \frac{c + 2b\sin A}{3} \qquad y'_3 = \frac{a - c + 2b\cos A}{3}$$

Since for an arbitrary x, $y = f(x) \in R(f)$, the angle A satisfies the cosine condition and thus $y' = f(x')$ becomes

$$a' = \sqrt{(x'_3 - x'_2)^2 + (y'_3 - y'_2)^2} = b^2 + c^2 - 2\ b\ c\cos A = a$$

$$b' = \sqrt{(x'_3 - x'_1)^2 + (y'_3 - y'_1)^2} = b$$

$$c' = \sqrt{(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2} = c$$

Since $y' \in R(f)$, the angles A', B', C' are determined from $a' = a, b' = b, c' = c$ through the cosine conditions and are therefore the same as those in y, i.e., $A' = A, B' = B, C' = C$. Therefore $y' = y$ and g is in fact a generalized inverse of f.

Turning to the range $R(g)$ of the generalized inverse g: for an arbitrary $y \notin R(f)$, $x = g(y)$ is given from the defining equations of g which show that x satisfies at least one condition, namely $x_1 = x_2$. The remaining five equations depend on four parameters, a, b, c, A, so there must be one more condition which results from the easily derived relations $y_2 - y_1 = c$ and $x_1 + x_2 + x_3 = c$. Therefore $x \in R(g)$ if its elements satisfy

$$x_1 = x_2, \qquad x_1 + x_2 + x_3 = y_2 - y_1$$

Thus $R(g)$ has dimension 6–2 = 4. The parameters $q = [abcA]^T$ may serve as curvilinear coordinates for $R(g)$ since the relation $x = g(y)$ degenerates into $x = x(q)$.

We now consider the fibering $\mathscr{F}'$ induced by $\mathscr{F}$ on $R(g)$. The elements of $R(g)$ which satisfy $f(x) = y$ for a fixed $y \in R(f)$ must simultaneously satisfy the conditions for $x \in F_y$

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = a^2$$
$$(x_3 - x_1)^2 + (y_3 - y_1)^2 = b^2$$
$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = c^2$$

and the conditions for $x \in R(g)$

$$x_1 = x_2, \qquad x_1 + x_2 + x_3 = y_2 - y_1$$

The combination gives after some manipulation the five conditions

$$x_1 = x_2, \quad x_1 + x_2 + x_3 = c, \quad y_2 - y_1 = c$$

$$y_3 - y_1 = b\cos A, \qquad x_3 - x_1 = b\sin A$$

which are those satisfied by any $x \in F'_y = F_y \cap R(g)$. The dimension of F'_y is 6–5 = 1. The fibers F_y as y runs through $R(f)$ constitute a fibering $\mathscr{F}$ of X, while the fibers F'_y constitute the fibering $\mathscr{F}'$ induced by $\mathscr{F}$ on $R(g)$.

We now move on to the section $S = g(R(f))$. When $y \in R(f)$, A, B, C are determined from a, b, c and so too $x = g(y)$. From the definition of g and the fact that

$$\cos A = \frac{b^2 + c^2 - a^2}{2\ bc}, \quad \sin A = \frac{\sqrt{4b^2c^2 - (b^2 + c^2 - a^2)}}{2\ bc}$$

x becomes explicitly

$$x_1 = \frac{c}{3} - \frac{\sqrt{4b^2c^2 - (b^2 + c^2 - a^2)^2}}{6c}, \ y_1 = \frac{a}{3} - \frac{c}{2} - \frac{b^2 - a^2}{6c}$$

$$x_2 = \frac{c}{3} - \frac{\sqrt{4b^2c^2 - (b^2 + c^2 - a^2)^2}}{6c}, \ y_2 = \frac{a}{3} + \frac{c}{2} - \frac{b^2 - a^2}{6c}$$

$$x_3 = \frac{c}{3} + \frac{\sqrt{4b^2c^2 - (b^2 + c^2 - a^2)^2},}{3c}, \ y_3 = \frac{a}{3} + \frac{b^2 - a^2}{3c}$$

The elements of $x \in S$ are functions of three parameters a, b, c. Therefore S has dimension three and it is described by $x = x(q)$ where $q = [a\ b\ c]^T$ are curvilinear coordinates on S. The section S can also be described by three conditions on the coordinates of its points. The obvious one is $x_1 = x_2$, the second results from $x_1 + x_2 + x_3 = c = y_2 - y_1$. The remaining one comes from $y_1 + y_2 + y_3 = a$. Thus $x \in S$ when

$$x_1 = x_2, \qquad x_1 + x_2 + x_3 = y_2 - y_1$$

$$(y_1 + y_2 + y_3)^2 = (x_3 - x_2)^2 + (y_3 - y_2)^2$$

The first two conditions are those satisfied by the elements of range $R(g)$ and therefore $S = g(R(f)) \subset R(g) = g(Y)$.

For the mapping $p = f \circ g$ and the fibers $\pi_{\mathscr{P}}(y)$: setting $y' = p(y) = f(g(y)) = f(x)$, where $x = g(y)$, and using the definitions of f and g we obtain the explicit representation of p:

$$a' = \sqrt{b^2 + c^2 - 2\,bc\cos A}$$
$$b' = b$$
$$c' = c$$
$$A' = A$$
$$\cos B' = \frac{c - b\cos A}{a'} = \frac{c - b\cos A}{\sqrt{b^2 + c^2 - 2\,bc\cos A}}$$
$$\cos C' = \frac{b - c\cos A}{a'} = \frac{b - c\cos A}{\sqrt{b^2 + c^2 - 2\,bc\cos A}}$$

If we replace $b = b', c = c', A = A'$ we obtain the three cosine conditions so that $y' \in R(f)$ and in fact $R(p) = R(f)$, since y' depends on three parameters b, c, A, and therefore $R(q)$ has the same dimension, three, as $R(f)$.

For a fixed element $y_0 = [a_0\ b_0\ c_0\ A_0\ B_0\ C_0]^{\mathrm{T}}$, the fiber $\pi_{\mathscr{P}}(y_0)$ induced by p through y_0 consists of all the elements y such that $p(y) = p(y_0) = y'$, which means that $y \in \pi_{\mathscr{P}}(y_0)$ when its elements satisfy the condition $p(y) = p(y_0)$, which in view of the preceding description of p takes the explicit form

$$\sqrt{b^2 + c^2 - 2\,bc\cos A} = \sqrt{b_0^2 + c_0^2 - 2\,b_0c_0\cos A_0}$$
$$b = b_0$$
$$c = c_0$$
$$A = A_0$$
$$\frac{c - b\cos A}{\sqrt{b^2 + c^2 - 2\,bc\cos A}} = \frac{c_0 - b_0\cos A_0}{\sqrt{b_0^2 + c_0^2 - 2\,b_0c_0\cos A_0}}$$
$$\frac{b - c\cos A}{\sqrt{b^2 + c^2 - 2\,bc\cos A}} = \frac{b_0 - c_0\cos A_0}{\sqrt{b_0^2 + c_0^2 - 2\,b_0c_0\cos A_0}}$$

Since the first one and the last two are a direct consequence of the remaining three, the only conditions needed so that $y \in \pi_{\mathscr{P}}(y_0)$ are

$$a = \text{any},\ b = b_0,\ c = c_0,\ A = A_0,\ B = \text{any},\ C = \text{any}$$

This means that $\pi_{\mathscr{P}}(y_0)$ has dimension $6 - 3 = 3$ and the three parameters a, B, C can serve as a system of three curvilinear coordinates on $\pi_{\mathscr{P}}(y_0)$, which assumes the direct representation $y = y(q) = y(q, y_0)$ with explicit form

$$a = a,\ b = b_0,\ c = c_0,\ A = A_0,\ B = B,\ C = C$$

For the mapping $q = g \circ f$ and the fibers $\pi_{\mathscr{Q}}(x)$ we set $x' = q(x) = g(f(x)) = g(y)$, where $y = f(x) \in R(f)$, and therefore the angles A, B, C are functions of the sides a, b, c. We need in addition to $\cos A = \frac{b^2+c^2-a^2}{2bc}$ the corresponding relation

$$\sin A = \sqrt{1 - \cos^2 A} = \frac{T}{bc}$$
$$\text{where } T^2 = b^2c^2 - \left(\frac{b^2 + c^2 - a^2}{2}\right)^2$$

Using the explicit form of $y = f(x)$ we evaluate

$$\frac{b^2 + c^2 - a^2}{2} = (x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)$$
$$T = (x_3 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_3 - y_1)$$

Using the explicit form for f and the relations for $\cos A$ and $\sin A$ just given we obtain the explicit form of $x' = f(y)$:

$$x'_1 = \frac{c}{3} - \frac{T}{3c}, \quad y'_1 = \frac{a - c}{3} - \frac{1}{3c}\left(\frac{b^2 + c^2 - a^2}{2}\right)$$
$$x'_2 = \frac{c}{3} - \frac{T}{3c}, \quad y'_2 = \frac{a + 2c}{3} - \frac{1}{3c}\left(\frac{b^2 + c^2 - a^2}{2}\right)$$
$$x'_3 = \frac{c}{3} + \frac{2T}{3c}, \quad y'_3 = \frac{a - c}{3} + \frac{2}{3c}\left(\frac{b^2 + c^2 - a^2}{2}\right)$$

which in terms of the elements of x become

$$(x'_3 - x'_2)^2 + (y'_3 - y'_2)^2 = a^2 = (x_3 - x_2)^2 + (y_3 - y_2)^2$$
$$(x'_3 - x'_1)^2 + (y'_3 - y'_1)^2 = b^2 = (x_3 - x_1)^2 + (y_3 - y_1)^2$$
$$(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 = c^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$$

Comparing these expressions to the identical ones for F_y, we conclude that if $x' \in \pi_Q(x)$ then also $x' \in F_{f(x)}$ and vice versa. Thus the fibers of $\mathscr{Q}$ are identical to the corresponding solution spaces.

In order to express $x' = q(x) \in \pi_Q(x)$ as a function of x we must replace a, c, T and $\frac{1}{2}(b^2 + c^2 - a^2)$ with their expressions in terms of x and thus obtain

$$x'_1 = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{3} - \frac{(x_3 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_3 - y_1)}{3\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$
$$y'_1 = \frac{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} - \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{3} - \frac{(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)}{3\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$
$$x'_2 = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{3} - \frac{(x_3 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_3 - y_1)}{3\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$
$$y'_2 = \frac{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} + 2\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{3} - \frac{(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)}{3\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

$$x_3' = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{3} + \frac{2(x_3 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_3 - y_1)}{3\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

$$y_3' = \frac{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} - \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{3} + \frac{2[(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)]}{3\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

From the relations giving x' in terms of a, b, c it is obvious that the range $R(q)$ of q is identical with the section S.

As regards the fibering $\mathscr{G}$ included by g: let $y_0 = [a_0\ b_0\ c_0\ A_0\ B_0\ C_0]^T$ be a fixed element and consider the fiber $\pi_{\mathscr{G}}(y_0)$ (through y_0 induced by g). For any other point $y \in \pi_{\mathscr{G}}(y_0)$ it must hold that $g(y) = g(y_0) = x$ and using the explicit form of g we arrive at the conditions

$$3x_1 = c - b\sin A = c_0 - b_0 \sin A_0$$
$$3y_1 = a - c - b\cos A = a_0 - c_0 - b_0 \cos A_0$$
$$3x_2 = c - b\sin A = c_0 - b_0 \sin A_0$$
$$3y_2 = a + 2c - b\cos A = a_0 + 2c_0 - b\cos A_0$$
$$3x_3 = c + 2b\sin A = c_0 + 2b_0 \sin A_0$$
$$3y_3 = a - c + 2b\cos A = a_0 - c_0 + 2b_0 \cos A_0$$

The condition for x_2 has been discarded because it is identical to that for x_1. The remaining are not independent due to the relation between $\cos A$ and $\sin A$. Thus

$$y_1 + y_2 + y_3 = a = a_0$$
$$(x_3 - x_1)^2 + (y_3 - y_1)^2 = b^2 = b_0^2 \quad \Rightarrow \quad b = b_0$$
$$2x_1 + x_3 = c = c_0$$

Using $a = a_0$, $b = b_0$ and $c = c_0$ in the above conditions yields $\cos A = \cos A_0$, $\sin A = \sin A_0$, so that $A = A_0$. Therefore, $y \in \pi_{\mathscr{G}}(y_0)$ for a given fixed y_0, when its elements satisfy

$$a = a_0,\ b = b_0,\ c = c_0,\ A = A_0,\ B = \text{any},\ C = \text{any}$$

The fiber $\pi_{\mathscr{G}}(y_0)$ has dimension 6 – 4 = 2. The elements $q = [B\ C]^T$ may serve as a system of curvilinear coordinates so that $\pi_{\mathscr{G}}(y_0)$ is described by $y = y(q) = y(q, y_0)$ or explicitly

$$a = a_0,\ b = b_0,\ c = c_0,\ A = A_0,\ B = B,\ C = C$$

Proposition 1. A generalized inverse g of f is uniquely defined if the following are specified:

The range $R(g) \subset X$ of the generalized inverse g.
The section $S = g(R(f)) \subset R(g)$ of the fibering $\mathscr{F}$ of X induced by f.
The fibering $\mathscr{P}$ to be induced by $p = f \circ g$, i.e., all the fibers P_y corresponding to every $y \in R(f)$.
The refinement of the fibering $\mathscr{P}$ by the fibering $\mathscr{G}$, i.e., for every fiber $P_y \in \mathscr{P}$ its own fibering with members G_x corresponding to all $x \in F_y$.
For a given section M of the fibering $\mathscr{G}$ and every $y \in R(f)$, a corresponding mapping

$\gamma_y : M \cap \pi_{\mathscr{P}}(y) \to R(g) \cap F_y$

which is subject to the following restriction: if $\pi_{\mathscr{G}}(y) \cap M = \{y_0\}$ then $\gamma_y(y_0) \in F_y \cap S = \{g(y)\}$.

Proof. Consider any element $z \in Y$. Then the fibering $\mathscr{P}$ determines a unique fiber $\pi_{\mathscr{P}}(z)$ containing z, while $y = p(z) \in R(f)$ is determined by $R(f) \cap \pi_{\mathscr{P}}(z) = \{y\}$ and $\pi_{\mathscr{P}}(y) = \pi_{\mathscr{P}}(z)$. Since y uniquely determines F_y, $g(y)$ is also uniquely determined by $F_y \cap S = \{g(y)\}$. The fibering $\mathscr{G}$ determines a unique fiber $\pi_{\mathscr{G}}(z) = G_{g(z)}$ containing z and furthermore $\pi_{\mathscr{G}}(z) = G_{g(z)} \subset \pi_{\mathscr{P}}(z) = \pi_{\mathscr{P}}(y)$ due to the fact that $\mathscr{G}$ is a refinement of $\mathscr{P}$. Since M is a section of $\mathscr{G}$ it holds that $\pi_{\mathscr{G}}(z)$ intersects M at a unique element $\pi_{\mathscr{G}}(z) \cap M = \{z_0\}$. The known mapping γ_y defines an element $\gamma_y(z_0) \in R(g)$ and we define $g(z) \equiv \gamma_y(z_0)$. It remains to show that the mapping g defined for every $z \in Y$ by this procedure is in fact a generalized inverse of f. For an arbitrary $x \in X$ consider $z = f(x)$. Since $z \in R(f)$ it follows that in this case $y = z$, and by the restriction imposed on γ_y for its action on elements of $R(f)$ it must hold that $g(z) \equiv \gamma_y(z) \in F_z \cap S$ implying that $g(z) \in F_z$ and therefore

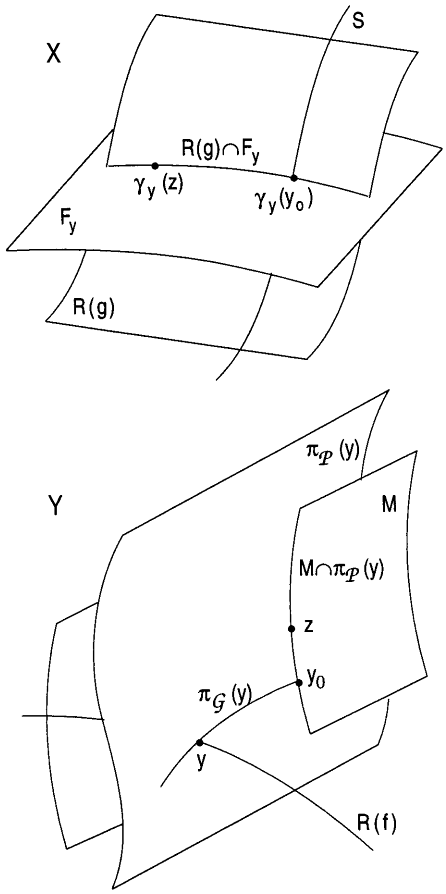

Fig. 4. An illustration of Proposition 1

$f(g(z)) = z$ or $f(g(f(x))) = (f \circ g \circ f)(x) = f(x)$ and since x has been arbitrarily chosen $f \circ g \circ f = f$ and g is indeed a generalized inverse of f. □

Example 1 (continued). We now consider the choice of a section M of $\mathscr{G}$. We have seen that $y \in \pi_{\mathscr{G}}(y_0)$ when

$$a = a_0,\ b = b_0,\ c = c_0,\ A = A_0,\ B = \text{any}, C = \text{any}$$

In order to create a section M of $\mathscr{G}$ we need only to pick up one element from each one of its fibers, and this can be done by assigning values to both B and C. If in addition we want $R(f) \subset M$, we must assign the known values

$$B_0 = \arccos\frac{a_0^2 + c_0^2 - b_0^2}{2a_0c_0} \quad \text{and} \quad C_0 = \arccos\frac{a_0^2 + b_0^2 - c_0^2}{2a_0b_0}$$

to B and C, respectively, whenever $y_0 \in R(f)$.

Thus M may be determined from the two conditions

$$\cos B = \frac{a^2 + c^2 - b^2}{2ac}, \qquad \cos C = \frac{a^2 + b^2 - c^2}{2ab}$$

and it has dimension 6–2 = 4. M contains the three-dimensional $R(f)$, which satisfies these conditions and in addition the similar one for $\cos A$. The independent parameters $a,\ b,\ c,\ A$ may serve as curvilinear coordinates for M which is then described by $y = y(q)$, explicitly

$$a = a,\ b = b,\ c = c,\ A = A$$

$$B = \arccos\frac{a^2 + c^2 - b^2}{2ac}, \qquad C = \arccos\frac{a^2 + b^2 - c^2}{2ab}$$

We now consider the intersection $M \cap \pi_{\mathscr{P}}(y_0) = \pi_{\mathscr{P}'}(y_0)$. If we combine the requirements for $y \in M$

$$a = \text{any},\ b = \text{any},\ c = \text{any},\ A = \text{any}$$

$$B = \arccos\frac{a^2 + c^2 - b^2}{2ac}, \qquad C = \arccos\frac{a^2 + b^2 - c^2}{2ab}$$

with those for $y \in \pi_{\mathscr{P}}(y_0)$

$$a = \text{any},\ b = b_0,\ c = c_0,\ A = A_0,\ B = \text{any},\ C = \text{any}$$

we conclude that $y \in M \cap \pi_{\mathscr{P}}(y_0) = \pi_{\mathscr{P}'}(y_0)$ when

$$a = \text{any},\ b = b_0,\ c = c_0,\ A = A_0$$

$$B = \arccos\frac{a^2 + c_0^2 - b_0^2}{2ac_0}, \qquad C = \arccos\frac{a^2 + b_0^2 - c_0^2}{2ab_0}$$

This means that $\pi_{\mathscr{P}'}(y_0) = M \cap \pi_{\mathscr{P}}(y_0)$ is of dimension 1, where the parameter a may serve as the single curvilinear coordinate. When, in particular, $y_0 \in R(f)$ then $A = A_0 = \arccos\frac{b_0^2 + c_0^2 - a_0^2}{2b_0c_0}$.

From the definition of g we see that when $y \in \pi_{\mathscr{P}'}(y_0)$ where $y_0 \in R(f)$ then $x = g(y) = \gamma_{y0}(y)$ is given by

$$x_1 = \frac{c_0 - b_0 \sin A_0}{3} \quad y_1 = \frac{a - c_0 - b_0 \cos A_0}{3}$$

$$x_2 = \frac{c_0 - b_0 \sin A_0}{3} \quad y_2 = \frac{a + 2c_0 - b_0 \cos A_0}{3}$$

$$x_3 = \frac{c_0 + 2b_0 \sin A_0}{3} \quad y_3 = \frac{a - c_0 + 2b_0 \cos A_0}{3}$$

Since $y_0 \in R(f)$ is fixed, the image $g(\pi_{\mathscr{P}'}(y_0))$ consists of elements x having $x_1,\ x_2,\ x_3$ constant, while $y_1,\ y_2,\ y_3$ depend on a single parameter a. It has therefore dimension 1. Since $x_0 = g(y_0)$ is fixed, $x = g(y)$ may be also expressed in terms of x_0:

$$x_1 = x_{01}, \quad y_1 = y_{01} + \tfrac{a - a_0}{3}$$

$$x_2 = x_{02}, \quad y_2 = y_{02} + \tfrac{a - a_0}{3}$$

$$x_3 = x_{03}, \quad y_3 = y_{03} + \tfrac{a - a_0}{3}$$

or

$$x = x_0 + \frac{a - a_0}{3}[0\ 0\ 0\ 1\ 1\ 1]^{\mathrm{T}}$$

We can also express $g(\pi_{\mathscr{P}'}(y_0))$ by five conditions. One possible choice is

$$x_1 = x_2,\ x_1 + x_2 + x_3 = c_0,\ y_2 - y_1 = c_0$$

$$x_3 - x_2 = b_0 \sin A_0,\ y_3 - y_1 = b_0 \cos A_0$$

Comparison with the identical conditions for $x \in F'_{y_0}$ shows that in fact $g(\pi_{\mathscr{P}'}(y_0)) = F'_{y_0} = F_{y_0} \cap R(g)$ is an element of the fibering $\mathscr{F}'$ induced by $\mathscr{F}$ on $R(g)$.

Returning once again to the theory, we would like the generalized inverse g of a continuous mapping f to be itself continuous, and this means that the mappings γ_y cannot be independent. Loosely speaking, they must map neighboring elements of Y belonging to neighboring fibers of P into neighboring elements of X. Thus γ_y may be considered as the restriction of a continuous mapping $\gamma\colon R(f) \times Y \to X$, defined by $\gamma(y,z) = \gamma_y(z)$, which is continuous in both y and z.

We can overcome the need to use a family of mappings γ_y, one for each $y \in R(f)$ with the introduction of additional fiberings for Y and X and a single mapping between them. Let M be a section of the fibering $\mathscr{G}$. In order to define g on $\pi_{\mathscr{P}}(y)$ for every $y \in R(f)$ it is sufficient to define it on $M \cap \pi_{\mathscr{P}}(y)$ since the fibering $\mathscr{G}$, being a refinement of $\mathscr{P}$, naturally extends g on the whole of $\pi_{\mathscr{P}}(y)$: for any $z \in \pi_{\mathscr{P}}(y) \notin M$ there exists a unique fiber $\pi_{\mathscr{G}}(z) \subset \pi_{\mathscr{P}}(z)$ from $\mathscr{G}$ such that $\pi_{\mathscr{G}}(z) \cap M = \{z_0\} \subset M \cap \pi_{\mathscr{P}}(z)$ and $g(z) = g(z_0)$. Although any section M of G is appropriate, it is more convenient to choose M so that it contains $R(f)$. This is possible in view of the next lemma.

Lemma. For every $y \in R(f)$, it holds that $\pi_{\mathscr{G}}(y) \cap R(f) = \{y\}$.

Proof. Since $\mathscr{G}$ is a refinement of $\mathscr{P}$, $\pi_{\mathscr{G}}(y) \subset \pi_{\mathscr{P}}(y)$. Since $R(f)$ is a section of $\mathscr{P}$ it holds $\pi_{\mathscr{G}}(y) \cap R(f) = \{y\}$ which combined with $\pi_{\mathscr{G}}(y) \subset \pi_{\mathscr{P}}(y)$ yields the desired result. □

The fibering $\mathscr{P}$ induces a fibering $\mathscr{P}' = \mathscr{P}_{|M}$ on M with fibers $\pi_{\mathscr{P}_{|M}}(z) = M \cap \pi_{\mathscr{P}}(z)$. $R(f)$ is a section of $\mathscr{P}$ and since $R(f) \subset M$ by choice, it is also a section of $\mathscr{P}'$. It is possible to consider a fibering $\mathscr{R}$ of M containing $R(f)$ which is complementary to $\mathscr{P}'$. Thus every $z \in M$ is the unique element $\pi_{\mathscr{R}}(z) \cap \pi_{\mathscr{P}'}(z) = \{z\}$. The generalized inverse g must be a mapping whose restriction to M is a one-to-one mapping $g_{|M}$ from M to $R(g)$, since

$\dim M = \dim R(g) = r_g$. A corresponding choice of complementary fiberings of $R(g)$ is possible. One is provided by the fibering $\mathscr{F}$ which induces a fibering $\mathscr{F}' = \mathscr{F}_{|R(g)}$ on $R(g)$ with fibers $\pi_{\mathscr{F}_{|R(g)}}(x) = \pi_{\mathscr{F}}(x) \cap R(g)$. For g to be a generalized inverse it must hold that $g(z) \in F_{p(z)} \cap R(g)$ so that each fiber of $\mathscr{P}'$ is mapped into the corresponding element of $\mathscr{F}'$. Since $z \in M$ is determined from one fibering of $\mathscr{P}'$ and one fibering of $\mathscr{R}$, it remains to introduce a fibering $\mathscr{S}$ of $R(g)$ complementary to $\mathscr{F}'$, such that $g(z)$ is determined by the intersection of a fiber of $\mathscr{S}$ with the already specified fiber of $\mathscr{F}'$. More precisely, to the fiber $\pi_{\mathscr{P}'}(z) \in \mathscr{P}'$ in M corresponds the fiber $F_{p(z)} \cap R(g) \in \mathscr{F}'$ in $R(g)$, and it remains to assign to the fiber $\pi_{\mathscr{R}}(z) \in \mathscr{R}$ in M a fiber $S' \in \mathscr{S}$ in $R(g)$ so that $g(z) = [F_{p(z)} \cap R(g)] \cap S'$. In particular, S must contain the particular section S of $\mathscr{F}$ which must be assigned to $R(f) \in \mathscr{R}$.

Proposition 2. A generalized inverse g of f, with $\mathrm{rank}(g) > \mathrm{rank}(f)$ is uniquely defined if the following are specified:

A fibering $\mathscr{P}$ of Y having $R(f)$ as a section.
A refinement $\mathscr{G}$ of the fibering $\mathscr{P}$.
A section M of the fibering $\mathscr{G}$ containing $R(f)$.
A manifold $R(g)$ of X with dimension $r_g > r$ on which $\mathscr{F}$ induces a fibering $\mathscr{F}'$ of dimension $d = m - r = \dim X - \mathrm{rank}(f)$.
A fibering $\mathscr{S}$ of $R(g)$ complementary to $\mathscr{F}'$.
A fibering $\mathscr{R}$ of M complementary to the fibering $\mathscr{P}'$ induced by $\mathscr{P}$ on M, which contains $R(f)$.
A mapping $\Gamma: \mathscr{R} \to \mathscr{S}$.
The generalized g is defined on M by $\{g(z)\} = [F_{p(z)} \cap R(g)] \cap \Gamma(\pi_{\mathscr{R}}(z))$ where $p(z)$ is defined by $\{p(z)\} = \pi_{\mathscr{P}'}(z) \cap R(f) = \pi_{\mathscr{P}}(z) \cap R(f)$.

Proof. We must show that the defined mapping g is a generalized inverse of f. For an arbitrary $x \in X$, $y = f(x) \in R(f) \subset M$. Since $y \in R(f)$ we have that $p(y) = y$, $\pi_{\mathscr{R}}(y) = R(f)$ and $\Gamma(\pi_{\mathscr{R}}(y)) = \Gamma\ (R(f)) \equiv S \in \mathscr{S}$. Applying the above definition,

$$\{g(y)\} = [F_{p(z)} \cap R(g)] \cap \Gamma(\pi_{\mathscr{R}}(y))$$
$$= [F_y \cap R(g)] \cap \Gamma(R(f)) = [F_y \cap R(g)] \cap S \subset F_y$$

and consequently $f(g(y)) = y$. Since $y = f(x)$ it follows that $f(g(f(x))) = (f \circ g \circ f)(x) = f(x)$ and since x is arbitrary $f \circ g \circ f = f$. □

Example 1 (continued). We begin with the fibering $\mathscr{R}$ of M containing $R(f)$. A fibering $\mathscr{R}$ of M can be introduced by fixing the value of the parameter a along the fibers $\pi_{\mathscr{P}'}(y_0) = M \cap \pi_{\mathscr{P}}(y_0)$, where we can assume without loss of generality that $y_0 \in R(f)$. If we want $\mathscr{R}$ to include $R(f)$ the value of $a = a(y)$ assigned to $y \in \pi_{\mathscr{P}'}(y_0)$ must be such that $a(y_0) = a_0$. A simple choice is to set $a = \lambda + a_0$, in which case λ replaces a as coordinate along $\pi_{\mathscr{P}'}(y_0)$, while $\lambda = 0$ gives the intersection y_0 of $\pi_{\mathscr{P}'}(y_0)$ with $R(f)$. With this choice a fiber R_λ from $\mathscr{R}$ is described by the elements y with

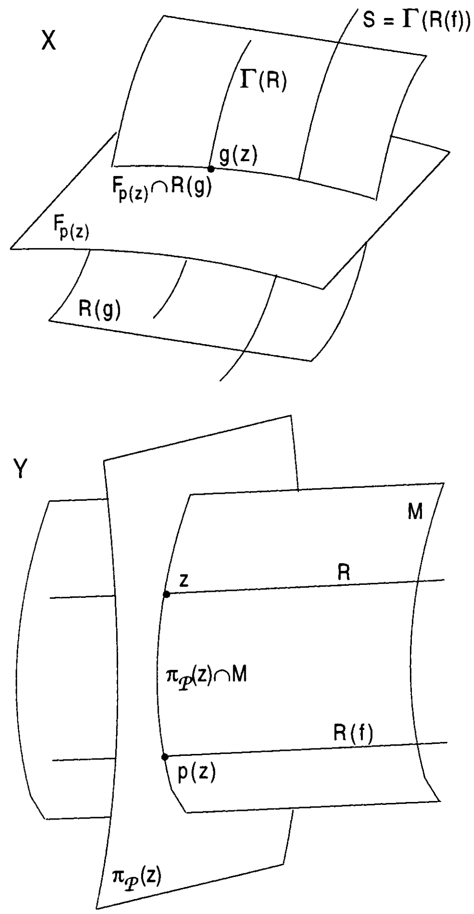

Fig. 5. An illustration of Proposition 2

$$a = \lambda + a_0, \quad b = b_0, \quad c = c_0, \quad A = A_0$$

$$B = \arccos\frac{a^2 + c^2 - b^2}{2ac}, \quad C = \arccos\frac{a^2 + b^2 - c^2}{2ab}$$

where $y_0 = p(y)$. From the definition of the projection p, we have that $y_0 = p(y)$ is given by

$$a_0 = \sqrt{b^2 + c^2 - 2bc\cos A}, \quad b_0 = b, \quad c_0 = c, \quad A_0 = A$$

$$B_0 = \arccos\frac{c - b\cos A}{\sqrt{b^2 + c^2 - 2bc\cos A}}$$

$$C_0 = \arccos\frac{b - c\cos A}{\sqrt{b^2 + c^2 - 2bc\cos A}}$$

Replacing these values in the preceding description of R_λ it follows that any $y \in R_\lambda$ satisfies

$$a = \lambda + \sqrt{b^2 + c^2 - 2bc\cos A}, \; b = b, \; c = c$$

$$A = A, \quad B = \arccos\frac{a^2 + c^2 - b^2}{2\ a\ c}, \quad C = \arccos\frac{a^2 + b^2 - c^2}{2\ a\ b}$$

R_λ is therefore of dimension 3, since it satisfies the $6 - 3 = 3$ conditions

$$(a - \lambda)^2 = b^2 + c^2 - 2\ b\ c\ \cos A$$

$$\cos B = \frac{a^2 + c^2 - b^2}{2\ a\ c}, \quad \cos C = \frac{a^2 + b^2 - c^2}{2\ a\ b}$$

In order to describe the image $S_\lambda \equiv g(R_\lambda)$ we need first

$$\cos A = \frac{b^2 + c^2 - (a - \lambda)^2}{2\ b\ c}$$

$$\Rightarrow \sin A = \frac{\sqrt{4b^2c^2 - \left[b^2 + c^2 - (a - \lambda)^2\right]}}{2\ b\ c}$$

which used in the definition of g yield

$$x_1 = \frac{c}{3} - \frac{\sqrt{4b^2c^2 - \left[b^2 + c^2 - (a - \lambda)^2\right]}}{6\ c}$$

$$y_1 = \frac{a}{3} - \frac{c}{2} - \frac{b^2 - (a - \lambda)^2}{6\ c}$$

$$x_2 = \frac{c}{3} - \frac{\sqrt{4b^2c^2 - \left[b^2 + c^2 - (a - \lambda)^2\right]}}{6\ c}$$

$$y_2 = \frac{a}{3} + \frac{c}{2} - \frac{b^2 - (a - \lambda)^2}{6\ c}$$

$$x_3 = \frac{c}{6} + \frac{2\sqrt{4b^2c^2 - \left[b^2 + c^2 - (a - \lambda)^2\right]}}{3\ c}$$

$$y_3 = \frac{a}{3} + \frac{b^2 - (a - \lambda)^2}{3\ c}$$

S_λ has dimension 3 since it is determined by three parameters a, b, c (λ is fixed). It can also be determined by three conditions, e.g.,

$$x_1 = x_2, \qquad x_1 + x_2 + x_3 = y_2 - y_1$$

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = (y_1 + y_2 + y_3 - \lambda)^2$$

Let $y \in R_\lambda \cap \pi_{\mathscr{P}'}(y_0)$. The y must satisfy simultaneously the conditions for $y \in R_\lambda$

$$(a - \lambda)^2 = b^2 + c^2 - 2\ b\ c \cos A$$

$$\cos B = \frac{a^2 + c^2 - b^2}{2\ a\ c}, \quad \cos C = \frac{a^2 + b^2 - c^2}{2\ a\ b}$$

and the conditions for $y \in \pi_{\mathscr{P}'}(y_0)$

$$b = b_0, \quad c = c_0, \quad A = A_0$$

$$\cos B = \frac{a^2 + c_0^2 - b_0^2}{2\ a\ c_0}, \quad \cos C = \frac{a^2 + b_0^2 - c_0^2}{2\ a\ b_0}$$

The combination gives the six conditions

$$a = a_0 + \lambda, \quad b = b_0, \quad c = c_0, \quad A = A_0$$

$$B = \arccos \frac{(a_0 + \lambda)^2 + c_0^2 - b_0^2}{2(a_0 + \lambda)c_0}$$

$$C = \arccos \frac{(a_0 + \lambda)^2 + b_0^2 - c_0^2}{2(a_0 + \lambda)b_0}$$

which uniquely determine y.

The image $g(y)$ of the y is given, according to the definition of g, by

$$x_1 = \frac{c_0 - b_0 \sin A_0}{3} \quad y_1 = \frac{a_0 + \lambda - c_0 - b_0 \cos A_0}{3}$$

$$x_2 = \frac{c_0 - b_0 \sin A_0}{3} \quad y_2 = \frac{a_0 + \lambda + 2c_0 - b_0 \cos A_0}{3}$$

$$x_3 = \frac{c_0 + 2b_0 \sin A_0}{3} \quad y_3 = \frac{a_0 + \lambda - c_0 + 2b_0 \cos A_0}{3}$$

Let $x \in S_\lambda \cap F'_{y_0} = g(R_\lambda) \cap g(\pi_{\mathscr{P}'}(y_0))$. It must simultaneously satisfy the conditions for $x \in S_\lambda$

$$x_1 = x_2, \quad x_1 + x_2 + x_3 = y_2 - y_1$$

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = (y_1 + y_2 + y_3 - \lambda)^2$$

and those for $x \in F'_{y_0}$

$$x_1 = x_2, \quad x_1 + x_2 + x_3 = c_0, \quad y_2 - y_1 = c_0$$

$$x_3 - x_2 = b_0 \sin A_0, \quad y_3 - y_1 = b_0 \cos A_0$$

The combination gives the six conditions

$$x_1 = x_2, \quad x_1 + x_2 + x_3 = c_0, \quad y_2 - y_1 = c_0,$$

$$x_3 - x_2 = b_0 \sin A_0, \quad y_3 - y_1 = b_0 \cos A_0,$$

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = (y_1 + y_2 + y_3 - \lambda)^2$$

which uniquely determine x. Indeed solving this system we get

$$x_1 = \frac{c_0 - b_0 \sin A_0}{3} = x_2, \quad x_3 = \frac{c_0 + 2b_0 \sin A_0}{3}$$

$$y_1 = \frac{a_0 + \lambda - c_0 - b_0 \cos A_0}{3},$$

$$y_2 = \frac{a_0 + \lambda + 2c_0 - b_0 \cos A_0}{3},$$

$$y_3 = \frac{a_0 + \lambda - c_0 + 2b_0 \cos A_0}{3}$$

But the above components of x determined by $\{x\} = S_\lambda \cap F'_{y_0}$ are in fact identical to those of the image $x = g(y)$ where y is determined by $\{y\} = R_\lambda \cap \pi_{\mathscr{P}}'(y_0)$. Therefore g is uniquely determined from the mapping $\Gamma\colon R_\lambda \to S_\lambda$, which to each fiber $R_\lambda \in \mathscr{R}$ assigns a fiber $S_\lambda \in \mathscr{S}$, the correspondence having been established through the common parameter λ. The mapping $\Gamma\colon \mathscr{R} \to \mathscr{S}$, maps

$$R_\lambda = \Big\{ y | (a-\lambda)^2 = b^2 + c^2 - 2bc\cos A,$$
$$\cos B = \frac{a^2+c^2-b^2}{2\ a\ c},\ \cos C = \frac{a^2+b^2-c^2}{2\ a\ b} \Big\}$$

to

$$S_\lambda = \Big\{ x | x_1 = x_2,\ x_1 + x_2 + x_3 = y_2 - y_1,$$
$$(x_3-x_2)^2 + (y_3-y_2)^2 = (y_1+y_2+y_3-\lambda)^2 \Big\}$$

The other required correspondence between $\pi_{\mathscr{P}'}(y_0)$ and F'_{y_0} is naturally introduced through the common element y_0 used in the definitions of the fibers $\pi_{\mathscr{P}'}(y_0) \in \mathscr{P}'$ and the corresponding ones $F'_{y_0} \in \mathscr{F}'$, where $\mathscr{F}'$ is thus a fibering of $R(g)$ complementary to $\mathscr{S}$. For given $y_0 \in R(f)$ the fiber F'_{y_0} depends only on the mapping f and not on the generalized inverse mapping g.

Example 2. We shall now give an example where the generalized inverse is not defined directly but follows from the choice of fiberings and subspaces as well as a fibering mapping Γ, along the lines of Proposition 2.

We consider the same mapping f as in Example 1, but we will choose a different description of the range $R(f)$ and the solution spaces F_y, using this time c, A, B as the free parameters. In this case $y = [a\ b\ c\ A\ B\ C]^{\mathrm{T}} \in R(f)$ when

$$a = c\frac{\sin A}{\sin(A+B)},\quad b = c\frac{\sin B}{\sin(A+B)},\quad c = \text{any}$$

$$A = \text{any},\quad B = \text{any},\quad C = \pi - (A+B)$$

The conditions in this above description are in fact the well-known angle and sine conditions

$$A + B + C = \pi,\quad \frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

We will introduce a fibering $\mathscr{P}$ having $R(f)$ as a section by $y \in \pi_{\mathscr{P}}(y')$ when

$$a = \text{any},\ b = \text{any},\ c = c',\ A = A',\ B = B',\ C = \text{any}$$

If $y \in \pi_{\mathscr{P}}(y') \cap R(f)$ then y must satisfy simultaneously the conditions for $y \in \pi_{\mathscr{P}}(y')$:

$$c = c',\quad A = A',\quad B = B'$$

and the conditions for $y \in R(f)$:

$$a = c\frac{\sin A}{\sin(A+B)},\quad b = c\frac{\sin B}{\sin(A+B)},\quad C = \pi - (A+B)$$

which, combined, show that y is uniquely determined from

$$a = c'\frac{\sin A'}{\sin(A'+B')},\quad b = c'\frac{\sin B'}{\sin(A'+B')},\quad c = c'$$

$$A = A',\quad B = B',\quad C = \pi - (A'+B')$$

This means that $\pi_{\mathscr{P}}(y') \cap R(f) = \{y\}$ and $R(f)$ is indeed a section of $\mathscr{P}$, as required. We can therefore describe the fibers of $\mathscr{P}$ using $y_0 \in R(f)$, in which case $y \in \pi_{\mathscr{P}}(y_0)$ when

$$a = \text{any},\ b = \text{any},\ c = c_0,\ A = A_0,\ B = B_0,\ C = \text{any}$$

From the preceding conditions it follows that for an arbitrary y its projection $y_0 = p(y) = f(g(y))$ is given by

$$a_0 = c\frac{\sin A}{\sin(A+B)},\quad b_0 = c\frac{\sin B}{\sin(A+B)},\quad c_0 = c$$

$$A_0 = A,\quad B_0 = B,\quad C_0 = \pi - (A+B)$$

The next step is to introduce a refinement $\mathscr{G}$ of $\mathscr{P}$ such that each fiber of $\mathscr{G}$ intersects $R(f)$ at most one element. Our choice is $y \in \pi_{\mathscr{G}}(y')$ when

$$a = a',\quad b = b',\quad c = c',\quad A = A',\quad B = B',\quad C = \text{any}$$

Obviously $y \in \pi_{\mathscr{G}}(y') \Rightarrow y \in \pi_{\mathscr{P}}(y')$ and therefore $\pi_{\mathscr{G}}(y') \subset \pi_{\mathscr{P}}(y')$, so that $\mathscr{G}$ is indeed a refinement of $\mathscr{P}$.

If $y' \notin R(f)$ then obviously for any $y \in \pi_{\mathscr{G}}(y')$ also $y \notin R(f)$, i.e. $\pi_{\mathscr{G}}(y') \cap R(f) = \varnothing$. If $y' = y_0 \in R(f)$ then

$$a_0 = c_0\frac{\sin A_0}{\sin(A_0+B_0)},\quad b_0 = c_0\frac{\sin B_0}{\sin(A_0+B_0)},$$
$$C_0 = \pi - (A_0+B_0)$$

and $y \in \pi_{\mathscr{G}}(y_0)$ when

$$a = c_0\frac{\sin A_0}{\sin(A_0+B_0)},\quad b = c_0\frac{\sin B_0}{\sin(A_0+B_0)},\quad c = c_0$$

$$A = A_0,\quad B = B_0,\quad C = \text{any}$$

in which case $\pi_{\mathscr{G}}(y_0) \cap R(f) = \{y_0\}$ as required.

In order to choose a section M of $\mathscr{G}$ we must pick up a single element y from each fiber $\pi_{\mathscr{G}}(y')$. Since the elements a, b, c, A, B are already determined from the equal ones of y', we are left with the possibility of choosing one value of C for every fiber of $\mathscr{G}$. Our choice is $C = \pi - (A+B)$ so that $y \in M$ when

$$a = \text{any},\quad b = \text{any},\quad c = \text{any}$$

$$A = \text{any},\quad B = \text{any},\quad C = \pi - (A+B)$$

M is obviously a section of $\mathscr{G}$ since $y \in M \cap \pi_{\mathscr{G}}(y')$ is uniquely determined from

$$a = a',\ b = b',\ c = c',\ A = A',\ B = B',\ C = \pi - (A'+B')$$

The fibering $\mathscr{P}$ induces a fibering $\mathscr{P}'$ on M with fibers $\pi_{\mathscr{P}'}(y_0) = \pi_{\mathscr{P}}(y_0) \cap M$. Combining the conditions for $y \in M$ with those for $y \in \pi_{\mathscr{P}}(y_0)$ we conclude that $y \in \pi_{\mathscr{P}'}(y_0)$ when

$$a = \text{any},\ b = \text{any},\ c = c_0,\ A = A_0,\ B = B_0,$$
$$C = \pi - (A_0+B_0)$$

For the fibering $\mathscr{F}$ of the solution spaces $F_y(y \in R(f))$ we need a representation which uses c, A, B as free parameters: $x \in F_y$ whenever $f(x) = y$, i.e.,

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = c^2$$

$$\frac{(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} = \cos A$$

$$\frac{(x_3 - x_2)(x_2 - x_1) + (y_3 - y_2)(y_2 - y_1)}{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} = \cos B$$

For $R(g) = g(Y)$ we must choose a subset of X which has the same dimension 5 as M, since $R(g) = g(M)$ also. Our choice is to set $x \in R(g)$ when

$$y_2 = y_1$$

The fibering $\mathscr{F}'$ induced by $\mathscr{F}$ on $R(g)$ has fibers $F'_y = F_y \cap R(g)$. Thus taking $y_2 = y_1$ into account, the preceding conditions for $x \in F_y$ we obtain, after some algebraic manipulation, that $x \in F'_y$ when

$$y_2 = y_1,\ x_2 - x_1 = c,\ \frac{y_3 - y_1}{x_3 - x_1} = \tan A,\ \frac{y_3 - y_2}{x_3 - x_2} = \tan B$$

Since for every y the fiber $\pi_{\mathscr{P}'}(y)$ will be mapped onto the fiber $F'_{p(y)}$, it remains to introduce a fibering $\mathscr{R}$ of M which is complementary to $\mathscr{P}'$. We correspond to each pair of positive numbers α, β a fiber $R_{\alpha,\beta} \in \mathscr{R}$ which is defined by $y \in R_{\alpha,\beta}$ when

$$a = \alpha,\ b = \beta,\ c = \text{any},\ A = \text{any},\ B = \text{any},$$
$$C = \pi - (A + B)$$

In order to show that $\mathscr{R}$ is complementary to $\mathscr{P}'$ (within M), let $y \in R_{\alpha,\beta} \cap \pi_{\mathscr{P}'}(y_0)$. Combining the conditions for $y \in R_{\alpha,\beta}$ with the ones for $y \in \pi_{\mathscr{P}'}(y_0)$ we conclude that y is uniquely determined from

$$a = \alpha,\ b = \beta,\ c = c_0,\ A = A_0,\ B = B_0,$$
$$C = \pi - (A_0 + B_0)$$

Therefore $R_{\alpha,\beta} \cap \pi_{\mathscr{P}'}(y_0) = \{y\}$ and $\mathscr{R}$ is complementary to $\mathscr{P}'$.

One more step remains to complete the definition of the generalized inverse g of f, that is to introduce a mapping

$$\Gamma : \mathscr{R} \to \mathscr{S} : R_{\alpha,\beta} \to S_{\alpha,\beta}$$

where $\mathscr{S}$ is a fibering of $R(g)$ complementary to $\mathscr{F}'$.

We choose $S_{\alpha,\beta} = \Gamma(R_{\alpha,\beta})$ by requiring that $x \in S_{\alpha,\beta}$ when

$$x_1 = \alpha, \quad y_1 = \beta, \quad y_2 = \beta$$

Obviously $y_1 = y_2$ and thus $S_{\alpha,\beta} \subset R(g)$. If $x \in S_{\alpha,\beta} \cap F'_{y_0}$ then combining these conditions for $x \in S_{\alpha,\beta}$ with those for $x \in F'_{y_0}$ we obtain

$$x_1 = \alpha, \quad y_1 = \beta, \quad y_2 = \beta$$

$$x_2 - x_1 = c_0, \quad \frac{y_3 - y_1}{x_3 - x_1} = \tan A_0, \quad \frac{y_3 - y_2}{x_3 - x_2} = \tan B_0$$

These six conditions can be solved for the elements of x which is then uniquely determined by

$$x_1 = \alpha, \quad x_2 = \alpha + c_0, \quad y_1 = \beta, \quad y_2 = \beta$$

$$x_3 = \alpha + b_0 \cos A_0, \quad y_3 = \beta + b_0 \sin A_0$$

Since $S_{\alpha,\beta} \cap F'_{y_0} = \{x\}$, the fibering $\mathscr{S}$ is complementary to $\mathscr{F}'$ within $R(g)$.

In order to obtain finally an explicit description of the generalized inverse mapping $g : y \to x = g(y)$, let $y = [a\ b\ c\ A\ B\ C]^{\mathrm{T}}$ be an arbitrary element of Y. Then $y \in R_{\alpha,\beta} \cap \pi_{\mathscr{P}'}(y_0)$ when $y_0 = p(y)$ is given by

$$a_0 = c\frac{\sin A}{\sin(A + B)}, \quad b_0 = c\frac{\sin B}{\sin(A + B)}, \quad c_0 = c$$

$$A_0 = A, \quad B_0 = B, \quad C_0 = \pi - (A + B)$$

and $x = g(y)$ is uniquely determined from $x \in S_{\alpha,\beta} \cap F'_{y_0}$. This means that $x = g(y)$ if we replace $\alpha = a, \beta = b$ and the values of the elements of y_0 in the expression giving $x \in S_{\alpha,\beta} \cap F'_{y_0}$. This results in

$$x_1 = a, \quad x_2 = a + c, \quad y_1 = b, \quad y_2 = b$$

$$x_3 = \alpha + c\ \frac{\sin B}{\sin(A + B)}\cos A, \quad y_3 = \beta + c\frac{\sin B}{\sin(A + B)}\sin A$$

thus bringing to an end Example 2.

A special case of interest is the case of a *minimal rank* generalized inverse g where $r_g = r$. In this case

$$r_g = \dim(R(g)) = r = \dim(R(f))$$

which combined with $S \subset R(g)$ implies that $S = R(g)$.

Lemma. When $r_g = r$ the fibering $\mathscr{G}$ induced by g coincides with the fibering $\mathscr{P}$ induced by $p = f \circ g$.

Proof. Since $S = R(g)$, the restriction $g_{|R(f)}$ of g to the range of f is in this case a bijection between $R(f)$ and $R(g)$. For every $z \notin R(f)$, $g(z) \in R(g) = S$ and since S is a section of $\mathscr{F}$, $\{g(z)\} = S \cap F_y$ for a unique fiber F_y corresponding to a unique $y \in R(f)$. Therefore $y = f(g(z)) = (f \circ g)(z) = p(z)$ and $z \in \pi_{\mathscr{P}}(y) = \pi_{\mathscr{P}}((f \circ g)(z))$. Consequently the fibering $\mathscr{G}$ induced by g coincides with the fibering $\mathscr{P}$ induced by $p = f \circ g$. □

For the particular choice $r_g = r$, the last two requirements in Proposition 1 can be relaxed. The same simplified situation can arise in a different way by requiring that in addition to g being a generalized inverse of f, at the same time f is a generalized inverse of g. Repeating property (G1) by interchanging the roles of f and g leads to the following definition.

Definition. A generalized inverse g of a given mapping f is called a *reflexive* generalized inverse if in addition to (G1) it satisfies the following condition, henceforth denoted (G2):

$$g \circ f \circ g = g \qquad (8)$$

Lemma. A generalized inverse g of f is a reflexive generalized inverse of f if and only if $r_g = r$.

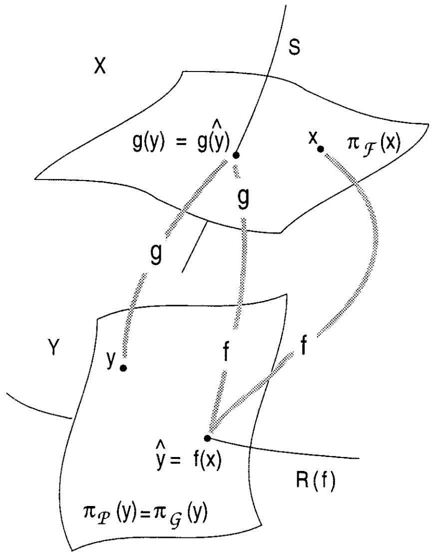

Fig. 6. The geometry of a reflexive generalized inverse g of a nonlinear mapping f. All elements of $\pi_{\mathscr{P}}(y) = \pi_{\mathscr{G}}(y)$ are mapped into the same element $\hat{x} \in S$ and thus projected by p onto the same element $\hat{y}$

Proof. Necessity: an immediate consequence of g being a generalized inverse of f is that $R(g \circ f) = R(g)$, and $r_g \leq \min(r, r_g)$, so that $r_g \leq r \leq \min(n, m)$, which combined with $r \leq r_g$ from Eq. (5) implies that $r_g = r$.

Sufficiency: if $r = r_g$ then $\dim(R(g)) = \dim(R(f)) = \dim(S)$ and $S = R(g)$. If $x \in S$ then $\{x\} = S \cap F_{f(x)}$ and $g(f(x)) = x$. Let y be an arbitrary element of Y, then $g(y) \in R(g) = S$ and setting $g(y)$ in the place of the previous x we obtain $g(f(g(y))) = g(y)$ or $(g \circ f \circ g)(y) = g(y)$ and since y is arbitrary $g \circ f \circ g = g$. □

Proposition 3. A reflexive generalized inverse g of f is uniquely defined if the following are specified:

(1) The section $S = g(R(f)) = R(g)$ of the fibering $\mathscr{F}$ of X induced by f.
(2) The fibering $\mathscr{P}$ to be induced by $p = f \circ g$, i.e., all the fibers P_y corresponding to every $y \in R(f)$.

The question of what the fibers G_x of a reflexive generalized inverse are, is answered by the following lemma.

Lemma. For a reflexive generalized inverse g, the $(f \circ g)$-induced fibers are identical to the g-induced fibers, i.e., $\mathscr{P} = \mathscr{G}$.

Proof. Consider any $y \in R(f)$. $z \in \pi_{\mathscr{P}}(y) \Rightarrow y = (f \circ g)(z) \Rightarrow g(y) = (g \circ f \circ g)(z) = g(z) \Rightarrow z \in \pi_{\mathscr{G}}(y) \Rightarrow \pi_{\mathscr{P}}(y) \subset \pi_{\mathscr{G}}(y)$. But since $\mathscr{G}$ is a refinement of $\mathscr{P}$, it must also hold that $\pi_{\mathscr{G}}(y) \subset \pi_{\mathscr{P}}(y)$ and thus $\pi_{\mathscr{G}}(y) = \pi_{\mathscr{P}}(y)$ for any $y \in R(f)$. Therefore $\mathscr{G} = \mathscr{P}$. □

Example 3. We shall give an example of a reflexive generalized inverse which is defined indirectly through the choice of the fibering $\mathscr{P} \equiv \mathscr{G}$ having $R(f)$ as a section and the section S of the fibering $\mathscr{F}$.

We choose $\mathscr{P}$ in such a way that for any given $y_0 \in R(f)$ an element $y \in \pi_{\mathscr{P}}(y_0)$ when

$$a_0 = a, \quad b_0 = a\frac{\cos\frac{A+C-B}{2}}{\cos\frac{B+C-A}{2}}, \quad c_0 = a\frac{\cos C}{\cos\frac{B+C-A}{2}}$$

$$A_0 = A + \frac{\pi - (A+B+C)}{2}, \; B_0 = B + \frac{\pi - (A+B+C)}{2}, \; C_0 = C$$

These equations describe also the projection $y_0 = p(y)$ for any given y.

For the corresponding fiber F_{y_0} of $\mathscr{F}$ we need the appropriate description $x \in F_{y_0}$ when

$$\frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\,\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} = -\cos B_0$$

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = a_0^2$$
$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = c_0^2$$

The definition of the generalized inverse g is finalized with the selection of the section S of $\mathscr{F}$ by $x \in S$ when

$$x_1 + x_2 + x_3 = 0, \quad y_1 + y_2 + y_3 = 0, \quad x_2 = x_3$$

For an arbitrary element y of Y the corresponding fiber $\pi_{\mathscr{P}}(y) = \pi_{\mathscr{P}}(y_0)$ is determined from the projection $y_0 = p(y)$. Since $R(g) = S$ while $\pi_{\mathscr{P}}(y_0)$ is mapped onto F_{y_0}, the image $x = g(y)$ must belong to both, $x \in S \cap F_{y_0}$. Combining the conditions for $x \in S$ and $x \in F_{y_0}$, we obtain a system of six equations in six unknowns

$$\frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\,\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} = -\cos B_0$$

$$(x_3 - x_2)^2 + (y_3 - y_2)^2 = a_0^2, \; (x_2 - x_1)^2 + (y_2 - y_1)^2 = c_0^2$$

$$x_1 + x_2 + x_3 = 0, \quad y_1 + y_2 + y_3 = 0, \quad x_2 = x_3$$

The solution is easily verified to be

$$x_1 = \tfrac{2}{3}c_0 \sin B_0, \qquad y_1 = -\tfrac{1}{3}a_0 + \tfrac{2}{3}c_0 \cos B_0$$

$$x_2 = -\tfrac{1}{3}c_0 \sin B_0, \qquad y_2 = -\tfrac{1}{3}a_0 - \tfrac{1}{3}c_0 \cos B_0$$

$$x_3 = -\tfrac{1}{3}c_0 \sin B_0, \qquad y_3 = \tfrac{2}{3}a_0 - \tfrac{1}{3}c_0 \cos B_0$$

These expressions give in fact not g but its restriction $g_{|R(f)}$ of g to R(f). This can be completed to $g = g_{|R(f)} \circ p$ from the known projection mapping p, by simply replacing B_0, a_0, c_0 with their expressions in terms of a, b, c, A, B, C. Thus we obtain the explicit form $x = g(y)$ of the reflexive generalized inverse g

$$x_1 = \frac{2}{3}\, a \cos C \frac{\cos\frac{A-B+C}{2}}{\cos\frac{B+C-A}{2}},$$

$$y_1 = -\tfrac{1}{3}a + \tfrac{2}{3}a \cos C \frac{\sin\frac{A-B+C}{2}}{\cos\frac{B+C-A}{2}}$$

$$x_2 = -\tfrac{1}{3}a \cos C \frac{\cos\frac{A-B+C}{2}}{\cos\frac{B+C-A}{2}},$$

$$y_2 = -\tfrac{1}{3}a - \tfrac{1}{3}a \cos C \frac{\sin\frac{A-B+C}{2}}{\cos\frac{B+C-A}{2}}$$

$$x_3 = -\tfrac{1}{3}a \cos C \frac{\cos\frac{A-B+C}{2}}{\cos\frac{B+C-A}{2}},$$

$$y_3 = \tfrac{2}{3}a - \tfrac{1}{3}a \cos C \frac{\sin\frac{A-B+C}{2}}{\cos\frac{B+C-A}{2}} \qquad \square$$

For reflexive generalized inverses we have a possibility of choices, corresponding to the choices of the section S of $\mathscr{F}$ and of the fibering $\mathscr{P}$ of Y over the elements of $R(f)$. In analogy to the uniquely defined pseudo-inverse of a linear operator we seek two conditions, additional to (G1) and (G2), such that (G3) specifies the fibering $\mathscr{P}$ and (G4) specifies the section S of $\mathscr{F}$.

We introduce the *minimum-distance generalized inverse*. Let ρ_Y be the distance function of the metric space Y. For any fixed element $\hat{y} \in R(f)$ define

$$R_{\hat{y}} = \{z \in Y; \rho_Y(z, \hat{y}) = \min_{y \in R(f)} \rho_Y(z, y)\} \tag{9}$$

Eq. (9) we name (G3). In the particular case that the subsets $R_{\hat{y}}$ are fibers of Y, as $\hat{y}$ varies over $R(f)$, then we can employ the following definition: g is a *minimum distance* generalized inverse of f, if $\mathscr{P} = \mathscr{R}$, i.e., when the fibering $\mathscr{P}$ induced by $p = f \circ g$ is identical with the fibering $\mathscr{R}$ with fibers $R_{\hat{y}}$.

If Y is an inner-product vector space with metric induced by a quadratic norm, we may use the term "*least squares*" in place of "minimum distance".

We now define the x_0-*nearest generalized inverse*. Let ρ_X be the distance function of $X, x_0 \in X$ be fixed and let

$$x_{F_y} = \min_{z \in F_y} \rho_X(x_0, z) \tag{10}$$

This equation we denote (G4). In the particular case that there exists only one element on every F_y which satisfies (G4), then the set of all x_{F_y} constitutes a section S of $\mathscr{F}$, and we can employ the following definition: g is an x_0-*nearest* generalized inverse of f if S is taken as the image of $R(f)$ under g, i.e., $S = g(R(f))$.

Remark. If X is an inner-product vector space with metric induced by the norm, while $\mathbf{x}_0 = \mathbf{0}$, we may use the term "minimum norm" in place of "**0**-nearest". However in the geodetic case **0** is a "prohibited" point, since it corresponds to a network with all its points coinciding.

Definition. The unique reflexive generalized inverse g of a given mapping f which is also a minimum distance and x_0-nearest generalized inverse is called the *pseudo-inverse* of f.

Various other nonunique generalized inverses can be defined by combination of (G1) with the some of the other properties (G2), (G3), and (G4). The inverses within the same class, i.e., the ones satisfying the same set of properties, may differ in the following aspects:

(1) In the section S when (G4) is not satisfied.
(2) In the fibering $\mathscr{P}$ induced by $p = f \circ g$ when (G3) is not satisfied.
(3) In the refinement $\mathscr{G}$ of $\mathscr{P}$ and the mapping $\gamma(y, z) = \gamma_y(z)$ extending g outside $R(f)$ when (G2) is not satisfied.

In the geodetic case property (G3) is a "must", since it solves the adjustment problem. Reflexivity (G2) is not necessary but convenient. The section S is specified either directly by a set of minimal constraints or by introducing (G4). (G4) is introduced, either directly (pseudo-inverse solution) or indirectly by a set of inner constraints which describe the particular section S corresponding to (G4).

Definition. A set of (nonlinear) equations $\chi(x, d) = 0$, is called a set of *minimal constraints* (with respect to f) if the corresponding mapping $\chi: X \times R^{m-r} \to R^{m-r}$ gives rise to a fibering $\mathscr{H}$ of X, such that its member $H_d = \{x \in H_d) | \chi(x, d) = 0\}$, corresponding to a fixed d, is a section of the fibering $\mathscr{F}$ induced by f.

Definition. A set of minimal constraints $h(x) = d$ are called *inner constraints* (with respect to a given point x_0 of X) when the fiber H_d coincides with the section of $\mathscr{F}$ specified by the property (G4), i.e., when for every $F_y \in \mathscr{F}$ the unique element x such that $F_y \cap H_d = \{x\}$ is the one closest to x_0 among all elements of F_y.

Remark. In the special case that f and its generalized inverse are linear or affine mappings, $R(f)$, the fibers F_y, P_y, G_x, Q_x, and the section S are all affine subspaces of the corresponding spaces Y and X. The fibers $\mathscr{F}, \mathscr{P}, \mathscr{G}, \mathscr{Q}$ consist of parallel affine subspaces, i.e., they are equivalent to quotient spaces of Y and X. Each quotient space is uniquely determined by its member passing through zero, i.e., by their "modulo" linear subspace in X or Y (Halmos 1974, Sect. 21) . The fiber $\mathscr{F}$, e.g., is equivalent to the quotient space of affine subspaces parallel to the null space of f. Thus the linear of affine generalized inverses are determined by a set of linear subspaces, instead of the fiberings (see Teunissen 1985). By the way, there is no reason to restrict the class of generalized inverses of a linear operator to be itself linear, as is usually done within the linear theory. A linear operator may well have a nonlinear generalized inverse. A geodetic example is the affine generalized inverse obtained implicitly by imposing a set of inhomogeneous linear constrains $\boldsymbol{Hx} = \boldsymbol{d}$, $\boldsymbol{d} \neq \mathbf{0}$, on the solution of the least-squares adjustment problem.

The approach followed here can be characterized as naive, and falls short of a proper mathematical theory of generalized inverses of nonlinear mappings, even in the relatively simple case of finite-dimensional spaces. The reason is that the assumptions made are too strong for the purpose of building which to build a general theory, a fact which will become obvious even in the simple geodetic case. Some of the assumptions that do not generally hold are the following:

(1) f may not be defined at all points of the space X but only on an open subset $U \subset X$.
(2) g may not be possible to define at all points of the space Y but only on an open subset $V \subset Y$.
(3) The minimum distance property (G3) may not suffice for the determination of a fibering of Y or even V. This has to do with the differential geometric properties of $R(f)$ as a submanifold of Y and in particular with its curvature tensor. There might be points in V such that their minimum distance from $R(f)$ does not correspond to a unique point of $R(f)$ but either on a set of discrete points or even a non discrete subset of $R(f)$. (Think, e.g., of the case where $R(f)$ is a ball in Y and consider its center which belongs to any fiber induced by the minimum distance principle.)
(4) The x_0-nearest property (G4) may not suffice for the determination of a fibering of X or even U. This has to do with the differential geometric properties of the fibers F_y as submanifolds of X. There might be one or more fibers F_y on which there is more than one point attaining the minimum distance from x_0.
(5) The rank of f (= rank of df_x) may not be the same at all points x of U. The consequence is that $R(f)$ is no longer a smooth submanifold of Y and its behavior at singular points (points with rank $(df_x) < r$) may pose additional problems.

We shall keep these problems in mind when studying the particular case of the mappings f arising in geodetic problems.

The last three problems have to do with points in Y or X which do not behave "nicely" with respect to the definitions employing (G3), (G4), and the definition of the rank r of f, $r = \text{rank}(df_x)$. Starting from the latter, we note that the set of points $y = f(x) \in R(f)$, for which $\text{rank}(df_x) < r$, forms within $R(f)$ a set of measure zero. This is in fact a consequence of Sard's theorem (Sternberg 1964, ch. II.3, Theorem 3.1), provided we impose some mild assumptions on the differentiability of the manifolds X and Y. In our case, where $\dim Y > \dim X$, it suffices to require that they are C^1 manifolds and f is also a mapping of class C^1.

A similar fact holds for the "problematic" points of Y which happen to belong to more than two sets $R_{\hat{y}}$ which thus intersect and fail to be fibers of Y. If such points form a subset $A = R_{\hat{y}_1} \cap R_{\hat{y}_2} \subset Y$ which is not of measure zero, then it will be possible to find a $y_0 \in A$ and a positive constant R such that $B_R(y_0) \subset A$, where $B_R(y_0)$ is the ball with center y_0 and radius R. We shall show that the assumption that A has nonzero measure leads to a contradiction: let y be a point on the geodesic joining y_0 with $\hat{y}_2$ such that $y \in B_R(y_0)$, in which case $\rho(y_0, \hat{y}_2) = \rho(y_0, y) + \rho(y, \hat{y}_2)$ and $\rho(y, \hat{y}_1) = \rho(y, \hat{y}_2)$ $\Rightarrow \rho(y_0, \hat{y}_1) = \rho(y_0, y) + \rho(y, \hat{y}_1)$, i.e., y belongs also to the geodesic joining y_0 and $\hat{y}_1$, which is obviously impossible.

For the definition based on (G4) we can ascertain that the set of problematic points forms a subset of X with measure zero, provided that the fibers $F_{f(x)}$ do not coincide locally with the boundaries of balls having x_0 as center.

In any case our definitions hold *almost everywhere* in X and Y, i.e., except on subsets of measure zero. From a physical point of view we can always work in a neighborhood of X and the corresponding neighborhood of Y which do not contain critical (problematic) points.

4 The geodetic nonlinear mapping

The nonlinear mapping f in the geodetic case arises from n individual mappings f_k, $k = 1, \dots n$, which map coordinates of points into various types of observables. We consider a geodetic network of N points with Cartesian coordinates $\boldsymbol{x}_i$, $i = 1, \dots, N$, where $\boldsymbol{x}_i$ is a vector of dimension either 2 (plane networks) or 3 (spatial networks). These vectors constitute a coordinate vector

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_N \end{bmatrix} \tag{11}$$

of dimension $2N$ or $3N$, which is considered an element of the m-dimensional Euclidean space ($m = 2N$ or $m = 3N$)

$$X = E^{2N} = \left(\times E^2\right)^N \text{ or } X = E^{3N} = \left(\times E^3\right)^N$$

equipped with the simple inner product $(\boldsymbol{x}_a, \boldsymbol{x}_b) = \boldsymbol{x}_a^{\mathrm{T}} \boldsymbol{x}_b$.

The geodetic observables we will examine here can be distinguished as "angular" observables y_{ijk} and "distance" observables d_{ik}, which differ with respect to their invariance characteristics under coordinate transformations. Angular observables are invariant under similarity coordinates transformations $\boldsymbol{x}' = S(\boldsymbol{x})$ while distance observables are invariant under rigid coordinate transformations $\boldsymbol{x}' = R(\boldsymbol{x})$, which are defined pointwise by

$$\boldsymbol{x}_i' = S(\boldsymbol{x}_i) = \lambda \, \boldsymbol{R} \, \boldsymbol{x}_i + \boldsymbol{t}, \quad \boldsymbol{x}_i' = R(\boldsymbol{x}_i) = \boldsymbol{R} \, \boldsymbol{x}_i + \boldsymbol{t} \tag{12}$$

$\boldsymbol{t}$ is a displacement vector, $\lambda > 0$ a scale parameter and $\boldsymbol{R}$ a proper orthogonal matrix ($\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{R}, \boldsymbol{R}^{\mathrm{T}} = \boldsymbol{I}$ and $|\boldsymbol{R}| = 1$) depending on a single rotational parameter θ in the two-dimensional case or on three rotational parameters $\theta_1, \theta_2, \theta_3$ in the three-dimensional case. The observables are functions of the coordinates

$$\begin{aligned} y_{ijk} &= y(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k) = y_{ijk}(\boldsymbol{x}) \\ d_{ik} &= d(\boldsymbol{x}_i, \boldsymbol{x}_k) = d_{ik}(\boldsymbol{x}) \end{aligned} \tag{13}$$

which have the invariance properties

$$y(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k) = y(S(\boldsymbol{x}_i), S(\boldsymbol{x}_j), S(\boldsymbol{x}_k))$$
$$d(\boldsymbol{x}_i, \boldsymbol{x}_k) = d(R(\boldsymbol{x}_i), R(\boldsymbol{x}_k)) \tag{14}$$

or

$$y_{ijk}(\boldsymbol{x}) = y_{ijk}(S(\boldsymbol{x})), \qquad d_{ik}(\boldsymbol{x}) = d_{ik}(R(\boldsymbol{x})) \tag{15}$$

The mapping f which maps the network coordinates $\boldsymbol{x}$ to the observables $\boldsymbol{y} = f(\boldsymbol{x}) \in R(f) \subset Y$ consists in general of n_1 angular mappings and n_2 distance mappings $(n_1 + n_2 = n)$ and we can distinguish two cases with respect to its invariance properties. When $n_2 = 0$ (only angular observations) then

$$f(\boldsymbol{x}) = f(S(\boldsymbol{x})) \tag{16}$$

while when $n_2 \neq 0$ (observations of distances or both distances and angles) then

$$f(\boldsymbol{x}) = f(R(\boldsymbol{x}))$$

Remark. We must emphasize here that f is a nonlinear mapping and S or R are general transformations not necessarily close to the identity (represented by $\lambda = 1, \boldsymbol{R} = \boldsymbol{I}, \boldsymbol{t} = \boldsymbol{0}$). It has been pointed out by Grafarend and Kampmann (1996) that there are other groups of transformations, such as the ten-parameter conformal group $G = C_{10}(3)$, in three dimensions; although it does not satisfy Eq. (16), i.e., $f(\boldsymbol{x}) \neq f(G(\boldsymbol{x}))$, when it is close to the identity ($G \approx \mathrm{Id}$), it satisfies the corresponding relation $\mathrm{d}f_{x_0}(\boldsymbol{x}) = \mathrm{d}f_{x_0}(G(\boldsymbol{x}))$ for the differential mapping $\mathrm{d}f_{x_0}$ of f at any Taylor point $\boldsymbol{x}_0$. In other words when G is close to the identity ("small" transformation) it leaves linearized angular observations invariant, a property shared of course with S. Therefore the $C_{10}(3)$ group, although of essentially no relevance to the strictly nonlinear approach followed here, should be taken into consideration in numerical applications, where we have to resort, explicitly or implicitly (iterative solutions), to linearization and solutions $\boldsymbol{x}, \boldsymbol{x}'$ close to $\boldsymbol{x}_0$, in which case the transformation from any such $\boldsymbol{x}$ to any such $\boldsymbol{x}'$ is necessarily close to the identity.

The presence of angular observations has as a consequence that f is not defined at every point of X. For an angular observable $y_{ijk} = y(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k)$ to be defined it is necessary that the relevant network points are distinct, i.e., we must exclude from X points $\boldsymbol{x}$ with $\boldsymbol{x}_i = \boldsymbol{x}_j$, or $\boldsymbol{x}_j = \boldsymbol{x}_k$, or $\boldsymbol{x}_k = \boldsymbol{x}_i$. Although the same problem does not appear in the case of distance observations, we will stick to the restriction that those network points which are joined by observations must be distinct. Let J be the index set of those pairs (i, k) of point indices corresponding to points joined by observations and let D_{ik} be the diagonal subspace of X defined by

$$D_{ik} = \{\boldsymbol{x} \in X | \boldsymbol{x}_i = \boldsymbol{x}_k\} \tag{18}$$

We must exclude all inappropriate diagonal subspaces, which leaves as the domain of definition of f the open subset $V \subset X$ defined by its complement V^{C} with respect to X

$$V^{\mathrm{C}} = \bigcup_{(i,k) \in J} D_{ik} = D \tag{19}$$

Since all D_{ik} are closed, the same holds for $D = V^{\mathrm{C}}$ and V is an open subset of X.

Turning to the range space Y we note that distances are mappings from X to R^+, since they obtain only positive values, while angles are mappings from X to the unit circle S. Thus, strictly speaking, f is a mapping

$$f : V \rightarrow (\times S)^{n_1} \times (\times R^+)^{n_2} \tag{20}$$

However it is standard practice to replace S with an interval of R, our choice being $I = (0, 2\pi]$, in which case f becomes a mapping

$$f : V \rightarrow U \equiv (\times I)^{n_1} \times (\times R^+)^{n_2} \subset Y = E^{n_1+n_2} = E^n \tag{21}$$

Since I does not contain its limit point 0 and the same holds true for R^+, U is an open subset of Y. We have assumed that $Y = E^n$ is R^n equipped with the euclidean inner product, though a more general inner product $(\boldsymbol{y}_a, \boldsymbol{y}_b) = \boldsymbol{y}_a^{\mathrm{T}} \boldsymbol{P} \boldsymbol{y}_b$ may be used, in which case $Y = E_P^n$.

The restriction of f from a mapping $f : X \rightarrow Y$ to a mapping $f : V \rightarrow U$, poses some serious problems as far as the properties (G3) and (G4) of generalized inverses are concerned. Now the range $R(f)$ is restricted from $f(X)$ to $U \cap f(X)$ and further to $U \cap f(V)$. The adjustment problem may have a solution $\hat{y} \in f(X)$ closest to a given $\boldsymbol{y} \in U$, such that $\hat{\boldsymbol{y}} \notin U \cap f(V)$. This means that adjusted observations $\hat{\boldsymbol{y}}$ may include negative or zero distances, or angles outside the prescribed interval.

Turning to the domain V we must distinguish between the similarity transformation and the rigid transformation case. When distances are observed and $\hat{\boldsymbol{y}} \in f(X)$, the fibers $F_{\hat{y}}$ may be initially defined with the help of any element $\boldsymbol{x} \in X$, such that $\hat{\boldsymbol{y}} = f(\boldsymbol{x}')$ by means of

$$F_{\hat{y}} = \{\boldsymbol{x} \in X | \boldsymbol{x} = R(\boldsymbol{x}')\} \tag{22}$$

This means that if $\boldsymbol{x}' \in V$ the same holds for any $\boldsymbol{x} = R(\boldsymbol{x}')$ and therefore $F_{\hat{y}} \subset V$. Thus $F_{\hat{y}}$ is either included in V or lies completely outside V. This means that the problem of finding an element $\hat{x}$ of a fiber closest to a given element $\boldsymbol{x}_0 \in V$ has a solution in V provided that $F_{\hat{y}} \subset V$, i.e., that $\hat{\boldsymbol{y}} \in f(V)$. When distance measurements are not included the similarity transformation S gives rise to fibers

$$F_{\hat{y}} = \{\boldsymbol{x} \in X | \boldsymbol{x} = S(\boldsymbol{x}')\} \tag{23}$$

which are no longer closed subsets of X, because they "converge" at the zero element $\boldsymbol{0} \in X$. Although the scale is restricted to positive values $\lambda > 0$, it is possible to consider a sequence $\lambda_i \rightarrow 0$, giving rise to a sequence of similarity transformations S_i such that $\boldsymbol{x}_i = S_i(\boldsymbol{x}') \in F_{\hat{y}} \rightarrow \boldsymbol{0} \notin F_{\hat{y}}$ (convergence in norm). As a consequence, the problem of finding an element $\hat{x}$ of a fiber closest to a given element, $\boldsymbol{x}_0 \in V$ may have no solution (consider the case where $\boldsymbol{0}$ is the closest to $\boldsymbol{x}_0$ element from the closure $\overline{F}_{\hat{y}}$ of the fiber $F_{\hat{y}}$).

These are problems of great concern to the mathematician, but of little or no concern at all to the geodesist. The reason is that in geodetic applications we are confined to a small neighborhood $N_{y*} \subset U \cap f(V)$ of the true observables $\boldsymbol{y}^* \in U \cap f(V)$, such that the observations $\boldsymbol{y}$ and the adjusted observations $\hat{y}$ also belong to $N_{\boldsymbol{y}^*}$. In the same way it is possible to choose an element $\boldsymbol{x}_0 \in V$ such that $F_{\hat{y}}$ crosses a small neighborhood $N_{\boldsymbol{x}_0} \subset V$ of $\boldsymbol{x}_0$ and furthermore the element $\hat{x}$ of $F_{\hat{y}}$ closest to $\boldsymbol{x}_0$ also lies in the same neighborhood $N_{\boldsymbol{x}_0}$.

Although we confine ourselves here to the deterministic aspects of generalized inverses, we are tempted to say that such a restriction to neighborhoods can be also justified when a probabilistic criterion is used for the "optimal" choice of a generalized inverse, which provides an estimator of x based on the observations y (outcomes of random variables). In this case a Bayesian point of view can be followed with a noninformative prior distribution for x, such that it is homogeneous within the desired neighborhood in X, and zero outside.

For a glimpse into the possible connection between the deterministic approach taken here and probabilistic aspects, consider the case where y is the outcome of a random variable Y with likelihood function $L(y,x) = p_Y(y|x)$. If it so happens that L has the form $L(x,y) = L(\rho(y,f(x)))$ where $L(\rho)$ has a maximum when ρ is minimized, while ρ qualifies as a metric, then maximum likelihood estimate can be identified with the solution following from the use of a minimum-distance generalized inverse.

With these considerations in mind we turn to the nonlinear datum problem. We assume that the adjustment problem has already been solved and its solution $\hat{\boldsymbol{y}}$ defines a specific fiber $F_{\hat{y}} \subset V \subset X$, described by either Eq. (22) or Eq. (23).

5 Solution to the nonlinear geodetic datum problem

The datum problem can be viewed as the problem of completing the projection mapping $p : Y \to R(f)$ which solved the adjustment problem into a (minimum-distance) generalized inverse $g = c \circ p$ of f, with the choice of a complementary mapping $c : R(f) \to S$, where S is a section of the fibering $\mathscr{F}$ induced by f in V. In fact it is sufficient to determine only S, in which case c (and thus g) is automatically defined as the mapping of any $\boldsymbol{y} \in R(f)$ into $c(\boldsymbol{y}) = \boldsymbol{x}$, where $\boldsymbol{x}$ is the unique element of $S \cap F_y$. One way of describing the d-dimensional manifold S is by means of a set of $d = m - r$ minimal constraints $\chi(\boldsymbol{x}, \boldsymbol{d}) = 0$, where $\boldsymbol{d}$ is a fixed vector of d parameters, which usually appear in the explicit form $h(\boldsymbol{x}) = \boldsymbol{d}$. The alternative possible description of S by the m constraints $\boldsymbol{x} = \phi(\boldsymbol{d}, \boldsymbol{q})$, where $\boldsymbol{d}$ is again a fixed vector of d parameters and $\boldsymbol{q}$ a vector of r free parameters, is hardly used in geodesy.

Strictly speaking, a set of minimal constraints consists of two things. First a family of mappings $\chi(\boldsymbol{x}, \cdot)$, or $\phi(\cdot, \boldsymbol{q})$, which corresponds to the choice of the type of constraints, and second a set of specific values for the parameters $\boldsymbol{d}$. For example, in a horizontal network we may decide to fix the two coordinates of a particular point and the azimuth of a particular side and thus choose the function $\chi(\boldsymbol{x}, \cdot)$ (or $h(\boldsymbol{x})$, or $\phi(\cdot, \boldsymbol{q})$, analogously). When we decide on the values to be assigned to the fixed coordinates and azimuth, we specify the value of the parameter set $\boldsymbol{d}$. The first step of this procedure corresponds to the choice of a fibering $\mathscr{S}$ complementary to $\mathscr{F}$, where each fiber corresponds to a particular set of values for $\boldsymbol{d}$. In the second step the assignment of specific values for $\boldsymbol{d}$ picks up a particular element S from $\mathscr{S}$ which will serve as the section of $\mathscr{F}$ into which g will map $R(f)$.

A particular choice for S is as the set of points $\boldsymbol{x}$ such that if $S \cap F_y = \{\boldsymbol{x}\}$ then $\boldsymbol{x}$ is the closest point to a given $\boldsymbol{x}_0 \in V$ among all elements of F_y. The description of this particular section by the corresponding set of minimal constraints (*inner constraints*) is not as easy as in the linear case. Instead we shall follow an approach similar to that of Baarda's S-transformation in the linear case, where a known reference "minimal" solution is transformed into the desired "inner" solution, i.e., the $\boldsymbol{x}_0$-nearest solution.

In order to proceed with the solution we need an analytical description of the solution fibers as submanifolds of X, e.g., by appropriate curvilinear coordinates.

Let G denote the group of applicable transformations (i.e., either similarity or rigid) with elements $g: X \to X$. If $\boldsymbol{z}$ belongs to a specific fiber $\boldsymbol{z} \in F_y$ then the same holds true for $\boldsymbol{x} = g(\boldsymbol{z}) \in F_y$ for every $g \in G$.

This establishes for every fiber F_y of X a natural correspondence $h : F_y \times F_y \to G$:

$$g = h(\boldsymbol{z}, \boldsymbol{x}) \Leftrightarrow \boldsymbol{x} = g(\boldsymbol{z}) \tag{24}$$

This means that by fixing one of the arguments of h we can obtain a mapping $h_z: \boldsymbol{x} \to h_z(\boldsymbol{x}) \equiv h(\boldsymbol{z}, \boldsymbol{x}): F_y \to G$ which is invertible. If further $f_G: G \to R^d$ is a coordinate system on G (i.e., a parametrization of G), the composite mapping

$$f \equiv f_G \circ h_z: \boldsymbol{x} \to \boldsymbol{p}(\boldsymbol{x}): F_y \to R^d \tag{25}$$

establishes a coordinate system on F_y. We have already introduced the coordinate systems f and the corresponding coordinates $\boldsymbol{p}$ in *the descriptions*, Eq. (12), of the similarity and rigid transformations. The dimension d of F_y depends on the applicable transformation group and the dimension of the network. For two-dimensional networks $d = 4$ for the similarity case (coordinates $\theta, t_1, t_2, \lambda$) and $d = 3$ for the rigid case (coordinates θ, t_1, t_2). For three-dimensional networks $d = 7$ for the similarity case (coordinates $\theta_1, \theta_2, \theta_3, t_1, t_2, t_3, \lambda$) and $d = 6$ for the rigid case (coordinates $\theta_1, \theta_2, \theta_3, t_1, t_2, t_3$). In fact we have introduced not one but a family of coordinate systems depending on the "reference point" $\boldsymbol{z}$ chosen for any fiber F_y. This choice can be made by introducing a section Z of the fibering $\mathscr{F}$ of X, e.g., by the use of minimal constraints which are easy to handle in computations, e.g., of the form $x_k^i = 0$ (*trivial constraints*), where k is the network point index and i the coordinate index ($i = 1, 2$ for plane and $i = 1, 2, 3$ for three-dimensional networks).

Once a reference solution $\boldsymbol{z}$ is established, all other solutions in the same fiber $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{z}, \boldsymbol{p})$ depend on the coordinates $\boldsymbol{p}$ which are identical to the parameters of the transformation $g(\boldsymbol{p}) : \boldsymbol{z} \to \boldsymbol{x}$. To obtain the solution specified by a given set of d minimal constraints $\chi(\boldsymbol{x}, \boldsymbol{d}) = 0$ (usually of the form $h(\boldsymbol{x}) = \boldsymbol{d}$) we have to solve a system of d nonlinear equations with d unknowns

$$\chi(\boldsymbol{p}) = \chi(\boldsymbol{x}(z, \boldsymbol{p}), \boldsymbol{d}) = \boldsymbol{0} \tag{26}$$

or in the usual form in geodesy

$$h(\boldsymbol{p}) = h(\boldsymbol{x}(z, \boldsymbol{p})) = \boldsymbol{d} \tag{27}$$

The solution $\boldsymbol{p}$ of the preceding system specifies the transformation $g_{\boldsymbol{p}} = g(\boldsymbol{p})$ which maps the reference solution z into the minimal constraint solution $\boldsymbol{x} = g_{\boldsymbol{p}}(z)$. This transformation is called the *nonlinear Baarda S-transformation* in the similarity, and the *nonlinear Baarda R-transformation* in the rigid case.

The solution $\boldsymbol{x}$ closest to a given element $\boldsymbol{x}_0 \in X$ is obtained either by minimizing

$$F(\boldsymbol{p}) = [\boldsymbol{x}(z, \boldsymbol{p}) - \boldsymbol{x}_0]^{\mathrm{T}}[\boldsymbol{x}(z, \boldsymbol{p}) - \boldsymbol{x}_0] \tag{28}$$

or by imposing the condition that the vector $\boldsymbol{x}_0 - \boldsymbol{x}$ is orthogonal to the tangent space to the fiber $F_{\hat{y}}$ at the point $\boldsymbol{x}$. Following the first approach we are led to a system of d nonlinear equations with d unknowns

$$\frac{\partial F}{\partial \boldsymbol{p}}(\boldsymbol{p}) = \boldsymbol{0} \Rightarrow \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}}\right)^{\mathrm{T}}(z, \boldsymbol{p})[\boldsymbol{x}(z, \boldsymbol{p}) - \boldsymbol{x}_0] = \boldsymbol{0} \tag{29}$$

The solution $\hat{\boldsymbol{p}}$ of this system specifies the transformation $g_{\hat{p}} = g(\hat{\boldsymbol{p}})$ which maps the reference solution z into the solution $\boldsymbol{x} = g_{\hat{p}}(z)$ closest to $\boldsymbol{x}_0$. $g_{\hat{p}}$ is the nonlinear Baarda S-transformation (or R-transformation) to the $\boldsymbol{x}_0$-nearest solution corresponding to the inner solution of Meissl in the linear case.

What about the famous inner constraints of Meissl which appear in the linear case? Well, $\frac{\partial F}{\partial p} = \boldsymbol{0}$ can be solved in principle to give $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{x}_0, z)$, which substituted in $\boldsymbol{x} = [g(\boldsymbol{p})](z) = \boldsymbol{x}(z, \boldsymbol{p})$ gives an expression of the form $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{x}_0, z) = \boldsymbol{x}(z)$, since $\boldsymbol{x}_0$ is fixed a priori. Recalling that z is determined by the intersection of a solution fiber $F_y \in \mathscr{F}$ with a section S_0 of $\mathscr{F}$ determined by a set of minimal constraints $\chi(z, \boldsymbol{d}) = 0$ or $z = \phi(\boldsymbol{d}, \boldsymbol{q})$, it is possible in principle to have a description of the r-dimensional manifold S_0 in the form $z = z(\boldsymbol{q})$, where $\boldsymbol{q}$ is a vector of r free parameters serving in fact as coordinates for S_0. If this representation is replaced in the inner solution $\boldsymbol{x} = \boldsymbol{x}(z)$ we obtain $\boldsymbol{x} = \boldsymbol{x}(z(\boldsymbol{q})) = \psi(\boldsymbol{q})$. The resulting relation $\boldsymbol{x} = \psi(\boldsymbol{q})$ is the set of inner constraints which to every $\boldsymbol{q}$ assigns the element of the fiber $\pi_{\mathscr{F}}(z(\boldsymbol{q}))$ of $\mathscr{F}$ passing through $z = z(\boldsymbol{q})$ its element closest to x_0. Unlike the linear case, the inner constraints are not independent of the reference minimal constraints because they depend on their parametrization in terms of $\boldsymbol{q}$. The above description of the inner constraints is a conceptual rather than a practical one since the analytical solution of nonlinear equations or the derivation of desired alternative descriptions of sections are not possible as a rule.

The detailed derivation of the specific form of the nonlinear system given by Eq. (29) is given in Appendix A for the four special cases corresponding to two- or three-dimensional networks and similarity or rigid transformations. The results are:

Three-dimensional networks – similarity transformation:

$$\boldsymbol{x}_i = \boldsymbol{x}_0 + \frac{\sum_i (\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)^{\mathrm{T}} \boldsymbol{R}(z_i - \bar{z})}{\sum_i (z_i - \bar{z})^{\mathrm{T}}(z_i - \bar{z})} \boldsymbol{R}(\boldsymbol{\theta})(z_i - \bar{z}) \tag{30}$$

where

$$\bar{z} \equiv \frac{1}{N}\sum_i z_i, \quad \bar{\boldsymbol{x}}_0 \equiv \frac{1}{N}\sum_i \boldsymbol{x}_{0i} \tag{31}$$

and $\boldsymbol{\theta}$ is the solution of the three nonlinear equations

$$\frac{1}{N}\sum_i [(\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)\times]\boldsymbol{R}(\boldsymbol{\theta})(z_i - \bar{z}) = \boldsymbol{0} \tag{32}$$

Three-dimensional networks – rigid transformation:

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_0 + \boldsymbol{R}(\boldsymbol{\theta})(z_i - \bar{z}) \tag{33}$$

where $\boldsymbol{\theta}$ is the solution of the three nonlinear equations

$$\frac{1}{N}\sum_i [(\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)\times]\boldsymbol{R}(\boldsymbol{\theta})(z_i - \bar{z}) = \boldsymbol{0} \tag{34}$$

Two-dimensional networks – similarity transformation:

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_0 + \frac{\sum_i (\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)^{\mathrm{T}} \boldsymbol{R}(z_i - \bar{z})}{\sum_i (z_i - \bar{z})^{\mathrm{T}}(z_i - \bar{z})} \boldsymbol{R}(\vartheta)(z_i - \bar{z}) \tag{35}$$

where ϑ is the solution of the single nonlinear equation

$$\frac{1}{N}\sum_i (\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)^{\mathrm{T}} \frac{\partial \boldsymbol{R}}{\partial \vartheta}(z_i - \bar{z}) = 0 \tag{36}$$

When $\boldsymbol{R}(\vartheta) = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix}$

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_0 + \frac{1}{s_z^2}\begin{bmatrix} b & a \\ -a & b \end{bmatrix}(z_i - \bar{z}) \tag{37}$$

where

$$a \equiv \frac{1}{N}\sum_i [(\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)^{\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}(z_i - \bar{z}) \tag{38}$$

$$b \equiv \frac{1}{N}\sum_i [(\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)^{\mathrm{T}}(z_i - \bar{z}) \tag{39}$$

$$s_z^2 \equiv \frac{1}{N}\sum_i [(z_i - \bar{z})^{\mathrm{T}}(z_i - \bar{z}) \tag{40}$$

Two-dimensional networks – rigid transformation:

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_0 + \boldsymbol{R}(\vartheta)(z_i - \bar{z}) \tag{41}$$

where ϑ is the solution of the single nonlinear equation

$$\frac{1}{N}\sum_i \left[(\boldsymbol{x}_{0i} - \bar{\boldsymbol{x}}_0)^{\mathrm{T}} \frac{\partial \boldsymbol{R}}{\partial \vartheta}(z_i - \bar{z})\right] = 0 \tag{42}$$

When $\boldsymbol{R}(\vartheta) = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix}$

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_0 \pm \frac{1}{\sqrt{a^2+b^2}} \begin{bmatrix} b & a \\ -a & b \end{bmatrix} (\boldsymbol{z}_i - \bar{\boldsymbol{z}}) \tag{43}$$

where a and b are defined as in the previous case.

6 Application to the linear case

We have outlined an approach to the representation of generalized inverses of nonlinear operators, and the question arises how this approach agrees with the existing representation theory for (linear) generalized inverses of linear operators (Takos 1976; Teunissen 1985). Since linear operators are simply special cases of nonlinear operators we will proceed to show that the existing theory can be derived as a special case of our nonlinear case.

If $f: X \to Y$ is a linear operator, we seek a linear operator $g: Y \to X$ which is a generalized inverse of f, i.e., it satisfies property (G1). In terms of given "initial" orthonormal bases $\{e_i^{X_0}\}$ for X and $\{e_i^{Y_0}\}$ for Y, f is represented by an $n \times m$ matrix $\boldsymbol{F}_0$ and we seek the $m \times n$ matrix $\boldsymbol{G}_0$ representing g. We call these bases initial because we are going to carry out our investigation in terms of "new" more convenient orthonormal bases $\{e_i^X\}$, $\{e_i^Y\}$, with respect to which f and g are represented by matrices $\boldsymbol{F}$ and $\boldsymbol{G}$, respectively. It is well known that if $\boldsymbol{U}$ and $\boldsymbol{V}$ are the orthogonal matrices of the change of bases, i.e.,

$$e_i^X = \sum_{k=1}^{m} U_i^k e_k^{X_0}, \qquad e_i^Y = \sum_{k=1}^{n} V_i^k e_k^{Y_0} \tag{44}$$

it holds that

$$\boldsymbol{F} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{U}, \quad \boldsymbol{G} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{G}_0\boldsymbol{V} \tag{45}$$

$$\boldsymbol{F}_0 = \boldsymbol{V}\,\boldsymbol{F}\,\boldsymbol{U}^{\mathrm{T}}, \quad \boldsymbol{G}_0 = \boldsymbol{U}\,\boldsymbol{G}\,\boldsymbol{V}^{\mathrm{T}} \tag{46}$$

For vectors $x \in X$ and $y \in Y$ it holds for their representations that

$$\boldsymbol{x} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_0, \quad \boldsymbol{y} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{y}_0 \tag{47}$$

$$\boldsymbol{x}_0 = \boldsymbol{U}\boldsymbol{x}, \quad \boldsymbol{y}_0 = \boldsymbol{V}\boldsymbol{y} \tag{48}$$

The choice of the new bases will be based on the so-called *singular value decomposition* (Rao and Mitra 1971, p.6) where $\{e_i^x\}$ and $\{e_i^y\}$ are the eigenvectors of the mappings $(f^* \circ f): X \to X$ and $(f \circ f^*): Y \to Y$, respectively, where $f^*: Y \to X$ is the *adjoint* mapping of f. Recall that if $(,)_X$ and $(,)_Y$ are the inner products in X and Y then f^* is defined by $(y, f(x))_Y = (f^*(y), x)_X$ for every $x \in X$ and $y \in Y$ (see, e.g., Halmos 1974, Sect. 44 and Sect. 68). In terms of the initial bases f^* is represented by $\boldsymbol{F}_0^{\mathrm{T}}$, $(f^* \circ f)$ by $\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{F}_0$, $(f \circ f^*)$ by $\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}$ while the eigenvector problems are expressed by

$$\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{u}_i = \lambda_i^2\boldsymbol{u}_i \Rightarrow \boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{F}_0 = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}^{\mathrm{T}} \tag{49}$$
$$\boldsymbol{U} = [\boldsymbol{u}_1\boldsymbol{u}_2\ldots\boldsymbol{u}_m], \qquad \boldsymbol{L}_{ik} = \delta_{ik}\lambda_i^2$$

$$\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{v}_i = \mu_i^2\boldsymbol{v}_i \Rightarrow \boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{V}\boldsymbol{M}\boldsymbol{V}^{\mathrm{T}} \tag{50}$$
$$\boldsymbol{V} = [\boldsymbol{v}_1\boldsymbol{v}_2\ldots\boldsymbol{v}_n], \qquad \boldsymbol{M}_{ik} = \delta_{ik}\mu_i^2$$

The eigenvectors $\{\boldsymbol{u}_i\}$, $\{\boldsymbol{v}_i\}$ form orthonormal systems and they can be seen as the representations with respect to the initial bases of the new bases, in which case $\boldsymbol{U}$ and $\boldsymbol{V}$ become the matrices of change of bases, already introduced with the same notation. Furthermore the number of non-zero eigenvalues is $r = \mathrm{rank}(f)$ for both λ_i^2 and μ_i^2, which are also identical, i.e., $\mu_i^2 = \lambda_i^2$, $i = 1, 2, \ldots r$. In this case, if $\boldsymbol{\Lambda}$ is the $r \times r$ diagonal matrix with elements

$$\Lambda_{ik} = \delta_{ik}\lambda_i \tag{51}$$

the representations with respect to the new bases are

$$\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} = \boldsymbol{U}^{\mathrm{T}}(\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{F}_0)\boldsymbol{U} = \boldsymbol{L} = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{52}$$

$$\boldsymbol{F}\boldsymbol{F}^{\mathrm{T}} = \boldsymbol{V}^{\mathrm{T}}(\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}})\boldsymbol{V} = \boldsymbol{M} = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{53}$$

$$\boldsymbol{F} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{U} = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{54}$$

The simplicity of this representation is the key to the specific choice of the new bases. This is essentially the same representation as in Takos (1976), since if the normality requirement is removed we can obtain a representation with respect to a new orthogonal (but not necessarily orthonormal) base

$$\boldsymbol{F}' = \begin{bmatrix} \boldsymbol{\Lambda}^{-1/2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-r} \end{bmatrix} \boldsymbol{V}^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{U} \begin{bmatrix} \boldsymbol{\Lambda}^{-1/2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{m-r} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{55}$$

We shall stick to the representation of Eq. (54) and will relate the section S, the fiberings $\mathscr{P}$, its refinement $\mathscr{G}$ which are needed for the determination g, to the submatrices of

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \tag{56}$$

where $\boldsymbol{T} = \boldsymbol{\Lambda}^{-1}$ follows directly from the property (G1): $\boldsymbol{F}\boldsymbol{G}\boldsymbol{F} = \boldsymbol{F}$.

From the discussion in Sect. 2 it is obvious that in the linear case the relevant fiberings consist of affine subspaces which are parallel translates of "generating" linear subspaces. They can be identified with the concept of quotient spaces. A *quotient space* of a linear space Z is a space $Z/M, M$ being a subspace of Z, consisting of equivalent classes of elements of Z, where two elements are equivalent, $\boldsymbol{x}_1 \sim \boldsymbol{x}_2$, if $\boldsymbol{x}_1 - \boldsymbol{x}_2 \in M$. The fibering $\mathscr{P} = Y/C$ is generated by a subspace C, its refinement $\mathscr{G} = Y/N(g)$ by $N(g)$ and the fibering $\mathscr{Q} = \mathscr{F} = X/N(f)$ by $N(f)$. The determination of g is completed by its action from C to the subspace $D = g(C)$.

Let us note first that the idempotent mappings $p = f \circ g$ and $q = g \circ f$ are in this case linear projections on $R(f)$ and S, respectively; p is represented by the matrix product $\boldsymbol{FG}$ and q by $\boldsymbol{GF}$, which explicitly are

$$\boldsymbol{FG} = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{\Lambda H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{57}$$

$$\boldsymbol{GF} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{J\Lambda} & \boldsymbol{0} \end{bmatrix} \tag{58}$$

Lemma. The subspace $R(f)$ is represented by $R\left(\begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0} \end{bmatrix}\right)$ while $N(f)$ is represented by $R\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix}\right)$.

Proof. $x \in N(f) \Leftrightarrow \boldsymbol{Gx} = \boldsymbol{0} \Leftrightarrow$

$$\begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda x}_1 \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{0} \;\Leftrightarrow\; \boldsymbol{x}_1 = \boldsymbol{0}$$

$$\Leftrightarrow \boldsymbol{x} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix} \boldsymbol{x}_2 \;\Leftrightarrow\; \boldsymbol{x} \in R\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix}\right).$$

$$y \in R(f) \Leftrightarrow \exists \quad \boldsymbol{x} \in R^m : \; \boldsymbol{y} = \boldsymbol{Fx} \Leftrightarrow$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda x}_1 \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{\Lambda x}_1 \in \mathbf{R}\left(\begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0} \end{bmatrix}\right)$$

□

Lemma. The section $S = g(R(f))$ is represented by the range $R\left(\begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{J\Lambda} \end{bmatrix}\right)$.

Proof. If $\hat{x} \in S = R(q)$ there exists $x \in X$ such that $\hat{x} = q(x)$ or

$$\hat{\boldsymbol{x}} = \boldsymbol{GFx} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{J\Lambda} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{J\Lambda x}_1 \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{J\Lambda} \end{bmatrix} \boldsymbol{x}_1 \in R\left(\begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{J\Lambda} \end{bmatrix}\right)$$

□

Lemma. The subspace $C \subset Y$ defined by $C = N(p) = N(f \circ g)$ is represented as the range $R\left(\begin{bmatrix} -\boldsymbol{\Lambda H} \\ \boldsymbol{I}_f \end{bmatrix}\right)$.

Proof. $y \in C \;\Leftrightarrow\; p(y) = (f \circ g)(y) = 0 \;\Leftrightarrow\; \boldsymbol{FGy} = \boldsymbol{0}$

$$\Leftrightarrow \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{\Lambda H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1 + \boldsymbol{\Lambda H y}_2 \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}$$

$$\Leftrightarrow \boldsymbol{y}_1 + \boldsymbol{\Lambda H y}_2 = \boldsymbol{0} \Leftrightarrow \boldsymbol{y}_1 = -\boldsymbol{\Lambda H y}_2 \Leftrightarrow$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Lambda H y}_2 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Lambda H} \\ \boldsymbol{I}_r \end{bmatrix} \boldsymbol{y}_2 \in R\left(\begin{bmatrix} -\boldsymbol{\Lambda H} \\ \boldsymbol{I}_r \end{bmatrix}\right)$$

□

The columns of $\begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{J\Lambda} \end{bmatrix}$ represent a basis for S, the columns of $\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix}$ a basis for $N(f)$, the columns of $\begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{0} \end{bmatrix}$ a basis for $R(f)$ and the columns of $\begin{bmatrix} -\boldsymbol{\Lambda H} \\ \boldsymbol{I}_f \end{bmatrix}$ a basis for C.

Since $X = S \oplus N(f)$ the columns of $\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{J\Lambda} \\ \boldsymbol{0} & \boldsymbol{I}_d \end{bmatrix}$ will represent a basis for X and we may set $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{J\Lambda} \\ \boldsymbol{0} & \boldsymbol{I}_d \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}$.

Since $Y = R(f) \oplus C$, the columns of $\begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{\Lambda H} \\ \boldsymbol{0} & \boldsymbol{I}_f \end{bmatrix}$ will represent a basis for Y and we may set $y = \begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{\Lambda H} \\ 0 & \boldsymbol{I}_f \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$.

Upon replacing these representations in $\boldsymbol{x} = \boldsymbol{Gy}$ we obtain

$$\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{J\Lambda} \\ \boldsymbol{0} & \boldsymbol{I}_d \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \boldsymbol{G} \begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{\Lambda H} \\ \boldsymbol{0} & \boldsymbol{I}_f \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{J\Lambda} \\ \boldsymbol{0} & \boldsymbol{I}_d \end{bmatrix}^{-1} \boldsymbol{G} \begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{\Lambda H} \\ \boldsymbol{0} & \boldsymbol{I}_f \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ -\boldsymbol{J\Lambda} & \boldsymbol{I}_d \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_r & 0 \\ \boldsymbol{0} & \boldsymbol{I}_f \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$$

and after carrying out the computations

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \mathcal{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \tag{59}$$

where

$$Q \equiv \boldsymbol{K} - \boldsymbol{J\Lambda H} \tag{60}$$

From the explicit form of Eq. (59) $\boldsymbol{x}_1 = \boldsymbol{\Lambda}^{-1}\boldsymbol{y}_1, \boldsymbol{x}_2 = \boldsymbol{Qy}_2$ it follows that the matrix $\boldsymbol{Q}$ determines the action of g on C, since if $y \in C$ then $\boldsymbol{y}_2 = \boldsymbol{0}, \boldsymbol{x}_1 = \boldsymbol{0}$ and

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{Qy}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix} \boldsymbol{Qy}_2 \in R\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix}\right)$$

The subspace $D = g(C)$ is represented by $R(\boldsymbol{D}), \boldsymbol{D}$ being any of the $m \times f$ matrices given by

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \begin{bmatrix} -\boldsymbol{\Lambda H} \\ \boldsymbol{I}_f \end{bmatrix} \boldsymbol{T} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{K} - \boldsymbol{J\Lambda H} \end{bmatrix} \boldsymbol{T} \tag{61}$$

where $\boldsymbol{T}$ is any $f \times f$ nonsingular matrix.

Summarizing these results we may say that the submatrix $\boldsymbol{J}$ determines the section S, the submatrix $\boldsymbol{H}$ determines the subspace $C = N(p)$ and thus the fibering $\mathcal{P}$ and the remaining submatrix $\boldsymbol{K}$ determines the subspace $D = g(C)$. The linear projection p represented by $\boldsymbol{FG}$ is a projection on $R(f)$, represented by $R(\boldsymbol{F})$, along the nullspace $N(p) = C$ represented by $R\left(\begin{bmatrix} -\boldsymbol{\Lambda H} \\ \boldsymbol{I}_f \end{bmatrix}\right)$. The linear projection q represented by $\boldsymbol{GF}$ is a projection on S represented by $R\left(\begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{J\Lambda} \end{bmatrix}\right)$ along the nullspace $N(q) = N(f)$ represented by $N(\boldsymbol{F})$.

We will now take a look into the special types of generalized inverses resulting by adding to the property

(G1) a combination from the properties (G2), (G3), (G4).

Lemma. For a *reflexive* generalized inverse $\boldsymbol{G}$ of $\boldsymbol{F}$ (g of f) it holds that $\boldsymbol{K} = \boldsymbol{J}\boldsymbol{\Lambda}\boldsymbol{H}$, i.e.,

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{J}\boldsymbol{\Lambda}\boldsymbol{H} \end{bmatrix} \tag{62}$$

Proof. It follows directly from the property (G2): $\boldsymbol{GFG} = \boldsymbol{G}$. □

In the linear case where X and Y are inner-product linear spaces with distances determined by the norm, the minimum-distance generalized inverse is called the *least-squares* generalized inverse and choosing $\boldsymbol{x}_0 = \boldsymbol{0}$, the $\boldsymbol{0}$-nearest generalized inverse is called *minimum- norm* generalized inverse.

Lemma. For a *least-squares* generalized inverse it holds that $\boldsymbol{H} = \boldsymbol{0}$, i.e.,

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{0} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \tag{63}$$

Proof. If $\boldsymbol{G}$ represents a least-squares inverse g then $\boldsymbol{FG}$ must represent an orthogonal projection p on $R(f)$. In other words, for any $y \in Y, \hat{y} = p(y)$ represented by $\hat{\boldsymbol{y}} = \boldsymbol{FGy}$ is the closest element to y from $R(f)$ if $\hat{y} - y \perp R(f)$, i.e., $\hat{y} - y \in R(f)^{\perp}$, where $R(f)^{\perp}$ is the orthogonal complement of $R(f)$ in X. Since $\hat{y} - y \in C$ and C is a subspace complementary to $R(f)$ it must hold that $C = R(f)^{\perp}$.

Since C is represented by $R\left(\begin{bmatrix} -\boldsymbol{\Lambda}\boldsymbol{H} \\ \boldsymbol{I}_f \end{bmatrix}\right)$ it must hold that

$$\begin{bmatrix} -\boldsymbol{\Lambda}\boldsymbol{H} \\ \boldsymbol{I}_f \end{bmatrix}^{\mathrm{T}} \boldsymbol{Fx} = \boldsymbol{0} \text{ for every } \boldsymbol{x}$$

$$\Leftrightarrow \begin{bmatrix} -\boldsymbol{\Lambda}\boldsymbol{H} \\ \boldsymbol{I}_f \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}^{\mathrm{T}} = \boldsymbol{0}$$

$$\Leftrightarrow -\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{x}_1 = \boldsymbol{0} \text{ for every } \boldsymbol{x}_1 \Leftrightarrow \boldsymbol{H} = \boldsymbol{0} \quad \square$$

Lemma. For a minimum-norm generalized inverse it holds that $\boldsymbol{J} = \boldsymbol{0}$, i.e.,

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{0} & \boldsymbol{K} \end{bmatrix} \tag{64}$$

Proof. If $\boldsymbol{G}$ represents a minimum-norm inverse g then q represented by $\boldsymbol{GF}$ must project every $x' \in F_y$ to an element $\hat{x} \in F_y$ which has the minimum norm among all elements of the same fiber F_y. Since $F_y = x' + N(f)$ this happens only when $\hat{x} \perp N(f)$. We have seen that q maps elements of X on the section S, and in view of the same dimension of these subspaces it must hold that $S = N(f)^{\perp}$. The columns of $\begin{bmatrix} \boldsymbol{\Lambda}^{-1} \\ \boldsymbol{J} \end{bmatrix}$ span S and must all be orthogonal to the columns of $\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix}$ spanning $N(f)$. This can be expressed in the form

$$\begin{bmatrix} \boldsymbol{\Lambda}^{-1} \\ \boldsymbol{J} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_d \end{bmatrix} = \boldsymbol{0} \Leftrightarrow \boldsymbol{\Lambda}_r^{-\mathrm{T}}\boldsymbol{0} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{I}_d = \boldsymbol{J}^{\mathrm{T}} = \boldsymbol{0}$$

$$\Leftrightarrow \boldsymbol{J} = \boldsymbol{0} \quad \square$$

Remark. Upon looking into the explicit form of the matrices $\boldsymbol{FG} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{\Lambda}\boldsymbol{H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$ and $\boldsymbol{GF} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{J}\boldsymbol{\Lambda} & \boldsymbol{0} \end{bmatrix}$, it is easy to see that $\boldsymbol{H} = \boldsymbol{0} \Leftrightarrow (\boldsymbol{FG})^{\mathrm{T}} = \boldsymbol{FG}$, while $\boldsymbol{J} = \boldsymbol{0} \Leftrightarrow (\boldsymbol{GF})^{\mathrm{T}} = \boldsymbol{GF}$. Thus we obtain the more familiar form of the properties (G3) and (G4) for least-squares and minimum-norm generalized inverses of matrices.

Corollary. A reflexive least-squares inverse is represented by

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{0} \\ \boldsymbol{J} & \boldsymbol{0} \end{bmatrix} \tag{65}$$

a reflexive minimum-norm inverse by

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{66}$$

a least squares-minimum norm inverse by

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K} \end{bmatrix} \tag{67}$$

and finally a reflexive least squares-minimum norm inverse (*pseudo inverse*) by

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{68}$$

□

A generalized inverse $\boldsymbol{G}$ of the matrix $\boldsymbol{F}$ is specified if the subspaces $C \subset Y$ and $S \subset X$ are known by means of specific representations and in addition the action of g on C is specified. Alternatively, $\boldsymbol{G}$ is uniquely determined once (in addition to C and S) we know the images under g of the elements of a basis of C, which images span and thus determine both the action of g on C and the subspace $D = g(C) \subset N(f) \subset X$.

Let C, S and D be represented in the original bases by $R(\boldsymbol{C}_0), R(\boldsymbol{S}_0)$ and $R(\boldsymbol{D}_0)$ with $\boldsymbol{D}_0 = \boldsymbol{G}_0\boldsymbol{C}_0$.

In the special bases introduced by the singular value decomposition C is represented by the range $R(\boldsymbol{C}) = R\left(\begin{bmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{bmatrix}\right)$ with $|\boldsymbol{C}_2| \neq 0$ (since $\dim C = f, \boldsymbol{C}$ is $n \times f$, so that $\boldsymbol{C}_1$ is $r \times f$ and $\boldsymbol{C}_2$ is $f \times f$), which means that the f columns of $\boldsymbol{C}$ represent the elements of a basis for C.

In the same special bases S is represented by the range $R(\boldsymbol{S}) = R\left(\begin{bmatrix} \boldsymbol{S}_1 \\ \boldsymbol{S}_2 \end{bmatrix}\right)$ with $|\boldsymbol{S}_1| \neq 0$ (since $\dim S = r, \boldsymbol{S}$

is $m \times r$, so that $\boldsymbol{S}_1$ is $r \times r$ and $\boldsymbol{S}_2$ is $d \times r$), which means that the r columns of $\boldsymbol{S}$ represent the elements of a basis for S.

The subspace $D = g(C)$ is represented by the matrix $\boldsymbol{D} = \boldsymbol{GC}$, which means that the f columns of $\boldsymbol{D}$ represent a spanning set of D, which, however, is not a basis of D because $\dim D = \Delta r < f$ in general.

In order to relate the two types of representations we may apply Eq. (48) to the columns of $\boldsymbol{S}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ to obtain

$$\boldsymbol{S}_0 = \boldsymbol{US}, \quad \boldsymbol{C}_0 = \boldsymbol{VC}, \quad \boldsymbol{D}_0 = \boldsymbol{UD} \tag{69}$$

with inverse relations

$$\boldsymbol{S} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}_0, \quad \boldsymbol{C} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{C}_0, \quad \boldsymbol{D} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}_0 \tag{70}$$

Setting

$$\boldsymbol{U} = [\boldsymbol{U}_1\boldsymbol{U}_2], \quad \boldsymbol{V} = [\boldsymbol{V}_1\boldsymbol{V}_2] \tag{71}$$

where $\boldsymbol{U}_1$ is $m \times r$, $\boldsymbol{U}_2$ is $m \times d$, $\boldsymbol{V}_1$ is $n \times r$ and $\boldsymbol{V}_2$ is $n \times f$, we obtain

$$\boldsymbol{S}_1 = \boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{S}_0, \qquad \boldsymbol{S}_2 = \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{S}_0 \tag{72}$$

$$\boldsymbol{D}_1 = \boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{D}_0, \qquad \boldsymbol{D}_2 = \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{D}_0 \tag{73}$$

$$\boldsymbol{C}_1 = \boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{C}_0, \qquad \boldsymbol{C}_2 = \boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{C}_0 \tag{74}$$

The relations can be used to obtain $\boldsymbol{G}_0$ (the representation of g in the original bases) in terms of the matrices $\boldsymbol{S}_0, \boldsymbol{C}_0, \boldsymbol{D}_0$ which represent the subspaces S, C, D, respectively, in the original bases.

Since $R\left(\begin{bmatrix}\boldsymbol{S}_1\\ \boldsymbol{S}_2\end{bmatrix}\right) = R\left(\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{J\Lambda}\end{bmatrix}\right)$, there exists an $r \times r$ non-singular matrix $\boldsymbol{L}$ such that $\begin{bmatrix}\boldsymbol{S}_1\\ \boldsymbol{S}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{J\Lambda}\end{bmatrix}\boldsymbol{L} = \begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{J\Lambda L}\end{bmatrix}$ which implies that $\boldsymbol{L} = \boldsymbol{S}_1$ and

$$\boldsymbol{J} = \boldsymbol{S}_2\boldsymbol{S}_1^{-1}\boldsymbol{\Lambda}^{-1} = \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{S}_0(\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{S}_0)^{-1}\boldsymbol{\Lambda}^{-1} \tag{75}$$

Since $R\left(\begin{bmatrix}\boldsymbol{C}_1\\ \boldsymbol{C}_2\end{bmatrix}\right) = R\left(\begin{bmatrix}-\boldsymbol{\Lambda H}\\ \boldsymbol{I}_f\end{bmatrix}\right)$ there exists an $f \times f$ non-singular matrix $\boldsymbol{T}$ such that $\left(\begin{bmatrix}\boldsymbol{C}_1\\ \boldsymbol{C}_2\end{bmatrix}\right) = \left(\begin{bmatrix}-\boldsymbol{\Lambda H}\\ \boldsymbol{I}_f\end{bmatrix}\right)\boldsymbol{T} = \begin{bmatrix}-\boldsymbol{\Lambda HT}\\ \boldsymbol{T}\end{bmatrix}$ which implies that $\boldsymbol{T} = \boldsymbol{C}_2$ and

$$\boldsymbol{H} = -\boldsymbol{\Lambda}^{-1}\boldsymbol{C}_1\boldsymbol{C}_2^{-1} = -\boldsymbol{\Lambda}^{-1}\boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{C}_0(\boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{C}_0)^{-1} \tag{76}$$

In correspondence to the specific representation of C by $R(\boldsymbol{C})$ the subspace $D = g(C)$ is represented by $R(\boldsymbol{D})$ where the $m \times f$ matrix $\boldsymbol{D}$ is now given by Eq. (61) with $\boldsymbol{T} = \boldsymbol{C}_2$

$$\boldsymbol{D} = \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{K} - \boldsymbol{J\Lambda H}\end{bmatrix}\boldsymbol{C}_2 = \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{QC}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{D}_2\end{bmatrix} \tag{77}$$

where

$$\boldsymbol{D}_1 = \boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{D}_0 = \boldsymbol{0} \tag{78}$$

$$\boldsymbol{D}_2 \equiv \boldsymbol{QC}_2 = (\boldsymbol{K} - \boldsymbol{J\Lambda H})\boldsymbol{C}_2 \tag{79}$$

and

$$\begin{aligned}\boldsymbol{K} = \boldsymbol{Q} + \boldsymbol{J\Lambda H} = \boldsymbol{D}_2\boldsymbol{C}_2^{-1} + \boldsymbol{J\Lambda H}\\ = \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{D}_0(\boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{C}_0)^{-1}\\ - \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{S}_0(\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{S}_0)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{C}_0(\boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{C}_0)^{-1}\end{aligned} \tag{80}$$

The upper zero submatrix $\boldsymbol{D}_1$ in $\boldsymbol{D}$ is a direct consequence of the fact the $D \subset N(f)$ where $N(f)$ is represented by $R\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}_{\mathrm{d}}\end{bmatrix}\right)$. The same fact is expressed in the original bases by the condition $\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{D}_0 = \boldsymbol{0}$.

Remark. The matrix $\boldsymbol{D}$ is more than a representation of the subspace D. This would be the case if the matrix $\boldsymbol{T}$ was chosen arbitrarily. The choice $\boldsymbol{T} = \boldsymbol{C}_2$ makes $\boldsymbol{D}$ equal to $\boldsymbol{GC}$ and thus the columns of $\boldsymbol{D}$ are images under g of the columns of $\boldsymbol{C}$, i.e., of the specific given description for C. Thus $\boldsymbol{D}$ not only describes the subspace D but it also describes the action of g on the subspace C. Due to the fact that $D \subset N(f)$ the relation $\boldsymbol{D} = \boldsymbol{GC}$ degenerates into $\boldsymbol{D}_2 = \boldsymbol{QC}_2$. Every column of $\boldsymbol{C}$ is uniquely determined by the corresponding column of $\boldsymbol{C}_2$. Indeed if $\boldsymbol{e}_i$ the i-th column of $\boldsymbol{I}_f$ then the i-th column of $\boldsymbol{C}$ is $\boldsymbol{Ce}_i = \begin{bmatrix}-\boldsymbol{\Lambda H}\\ \boldsymbol{I}_f\end{bmatrix}\boldsymbol{C}_2\boldsymbol{e}_i$ and is determined by $\boldsymbol{C}_2\boldsymbol{e}_i$, the i-th column of $\boldsymbol{C}_2$. Therefore $\boldsymbol{C}_2$ represents in a certain sense $\boldsymbol{C}$ and thus the subspace C. The matrix $\boldsymbol{Q}$, which is uniquely related to $\boldsymbol{K}$ once $\boldsymbol{H}$ and $\boldsymbol{J}$ have been fixed by the choice of $\boldsymbol{C}$ and $\boldsymbol{S}$, defines through the relation $\boldsymbol{D}_2 = \boldsymbol{QC}_2$ the action of g on C. □

Thus the submatrices $\boldsymbol{J}, \boldsymbol{H}$ and $\boldsymbol{K}$ which determine $\boldsymbol{G}$ have now been related to the matrices $\boldsymbol{C}_0, \boldsymbol{S}_0$, and $\boldsymbol{D}_0$ which determine the subspaces C, S, and D, respectively, as well as the action of g on C. The matrix $\boldsymbol{G}_0 = \boldsymbol{UGV}^{\mathrm{T}}$ can now be expressed in terms of the matrices $\boldsymbol{C}_0, \boldsymbol{S}_0$, and $\boldsymbol{D}_0$ and the change of bases matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, which also assume geometric representations.

Lemma. In the original basis $R(f)$ is represented by $R(\boldsymbol{V}_1)$, $R(f)^{\perp}$ by $R(\boldsymbol{V}_2)$, $N(f)$ by $R(\boldsymbol{U}_2)$ and $N(f)^{\perp}$ by $R(\boldsymbol{U}_1)$.

Proof. Since $\boldsymbol{V}$ and $\boldsymbol{U}$ are orthogonal we obtain

$$\boldsymbol{I} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} \Rightarrow$$

$$\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_1 = \boldsymbol{I}_r, \ \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_2 = \boldsymbol{I}_d, \ \boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_2 = \boldsymbol{0}, \ \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_1 = \boldsymbol{0}$$

$$\boldsymbol{I} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} \Rightarrow$$

$$\boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{V}_1 = \boldsymbol{I}_r, \ \boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{V}_2 = \boldsymbol{I}_d, \ \boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{V}_2 = \boldsymbol{0}, \ \boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{V}_1 = \boldsymbol{0}$$

$R(f)$ is represented in the special basis by $R\left(\begin{bmatrix}\boldsymbol{I}_r\\ \boldsymbol{0}\end{bmatrix}\right)$, so that in the original basis it will be represented according to Eq. (48) by the column range of $U\begin{bmatrix} I_r \\ 0 \end{bmatrix} = U_1$. From $U_2^T U_1 = 0$, it follows that $R(f)^\perp$ is represented by $R(U_2)$. $N(f)$ is represented in the special basis by $R\left(\begin{bmatrix} 0 \\ I_d \end{bmatrix}\right)$, so that in the original basis it will be represented according to Eq. (48) by the column range of $V\begin{bmatrix} 0 \\ I_d \end{bmatrix} = V_2$. From $V_1^T V_2 = 0$, it follows that $N(f)^\perp$ is represented by $R(V_1)$. □

Remark. Since the column space of a matrix remains the same when the matrix is multiplied from the left with a nonsingular matrix, we have the equally valid representations of $R(f)$ by $R(V_1\Lambda)$ and of $N(f)^\perp$ by $R(U_1\Lambda)$. □

Lemma.

$$G_0 = S_0[(U_1\Lambda)^T S_0]^{-1}[V_1^T - V_1^T C_0(V_2^T C_0)^{-1} V_2^T] + D_0[V_2^T C_0]^{-1} V_2^T \tag{81}$$

Proof.

$$\begin{aligned} G_0 = UGV^T &= [U_1 U_2]\begin{bmatrix} \Lambda^{-1} & H \\ J & K \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \\ &= U_1\Lambda^{-1}V_1^T + U_1 H V_2^T + U_2 J V_1^T + U_2 K V_2^T \\ &= U_1\Lambda^{-1}V_1^T - U_1\Lambda^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T \\ &\quad + U_2U_2^T S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T \\ &\quad + U_2U_2^T D_0(V_2^T C_0)^{-1}V_2^T \\ &\quad - U_2U_2^T S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T \\ &= U_1U_1^T S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T \\ &\quad - U_1U_1^T S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T \\ &\quad + U_2U_2^T S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T \\ &\quad + U_2U_2^T D_0(V_2^T C_0)^{-1}V_2^T \\ &\quad - U_2U_2^T S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T \end{aligned}$$

which in view of $I = UU^T = [U_1 U_2]\begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_1U_1^T + U_2U_2^T$ and the condition $U_1^T D_0 = 0$ becomes

$$\begin{aligned} G_0 &= S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T \\ &\quad - S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T \\ &\quad + (I - U_1U_1^T)D_0(V_2^T C_0)^{-1}V_2^T \\ &= S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T \\ &\quad - S_0(U_1^T S_0)^{-1}\Lambda^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T \\ &\quad + D_0(V_2^T C_0)^{-1}V_2^T \\ &= S_0[\Lambda U_1^T S_0]^{-1}[V_1^T - V_1^T C_0(V_2^T C_0)^{-1}V_2^T] \\ &\quad + D_0[V_2^T C_0]^{-1}V_2^T \end{aligned}$$

□

The representation of a general generalized inverse given by Eq. (81), despite its algebraic form and origin, has a purely geometric character. All the submatrices appearing in it have clear geometric meanings. A similar representation has been presented by Teunissen (1985, Eq. 2.8), which in our notation can be written as

$$G_0 = S_0[(C_0^\perp)^T F_0 S_0]^{-1}(C_0^\perp)^T + D_0[((F_0S_0)^\perp)^T C_0]^{-1}((F_0S_0)^\perp)^T \tag{82}$$

where for a $p \times q$ matrix M with $q < m$, $M^\perp$ denotes a $p \times (p-q)$ matrix such that $\text{rank}[M|M^\perp] = p$ and $M^T M^\perp = 0$ (The column span of $M^\perp$ is an orthogonal complement of the column span of M in E^p). We prove the identity of the two representations of Eqs. (81) and (82) in Appendix B.

In the representation of Eq. (82) the matrices $C_0^\perp$ and $(F_0S_0)^\perp$ are not uniquely defined, although G_0 is. Our (admittedly less elegant) representation by Eq. (81) gives a unique expression for G_0 in terms of the chosen specific descriptions of the subspaces C, S, D and the action of g on C, as well as, in terms of the specific descriptions of $R(f)$, $N(f)$, $R(f)^\perp$ and $N(f)^\perp$ associated with the nonzero eigenvalues and the eigenvectors of the matrices $F_0^T F_0$ and $F_0 F_0^T$. Of course the choice of descriptions for the above subspaces is not unique.

It is easy to derive corresponding representation for special types of generalized inverses. We can transform the relevant representations Eqs. (56), (62)–(68) in terms of the submatrices H, J, K to the original bases using Eqs. (75), (76), (80). An alternative approach is to transform the restricting properties $K = J\Lambda H$, $H = 0$, $J = 0$, to the original bases using Eqs. (75), (76), (80) and then apply the resulting conditions $D_0 = 0$, $V_1^T C_0 = 0$, $U_2^T S_0 = 0$, respectively, on the general representation Eq. (81). The results are:

Reflexive generalized inverse:

$$G_0 = S_0[(U_1\Lambda)^T S_0]^{-1}V_1^T[I - C_0(V_2^T C_0)^{-1}V_2^T] \tag{83}$$

Least-squares generalized inverse:

$$G_0 = S_0[(U_1\Lambda)^T S_0]^{-1}V_1^T + D_0[V_2^T C_0]^{-1}V_2^T] \tag{84}$$

Minimum-norm generalized inverse:

$$G_0 = U_1(V_1\Lambda)^{-1} - S_0[(U_1\Lambda)^T S_0]^{-1}V_1^T C_0(V_2^T C_0)^{-1}V_2^T + D_0[V_2^T C_0]^{-1}V_2^T \tag{85}$$

Reflexive least-squares inverse:

$$G_0 = S_0[(U_1\Lambda)^T S_0]^{-1}V_1^T \tag{86}$$

Reflexive-minimum norm inverse:

$$G_0 = S_0[(U_1\Lambda)^T S_0]^{-1}[V_1^T - V_1^T C_0(V_2^T C_0)^{-1}V_2^T] + D_0[V_2^T C_0]^{-1}V_2^T \tag{87}$$

Least squares-minimum norm inverse:

$$G_0 = U_1(V_1\Lambda)^{-1}[I - C_0(V_2^T C_0)^{-1}V_2^T] \tag{88}$$

Pseudo-inverse:

$$G_0 = U_1\Lambda^{-1}V_1^T = (U_1\Lambda)\Lambda^{-1}(V_1\Lambda)^{-1} \quad (89)$$

All these representations involve the submatrices C_0, S_0, V_1, V_2, $(V_1\Lambda)$ and $(U_1\Lambda)$ which have a clear geometrical meaning.

Appendix A

Derivation of the nonlinear Baarda transformations to the inner solution

We shall derive here the solution to the nonlinear Baarda transformations, which are coordinate transformations $x = S(z)$ from any minimum constraints solution z to the x_0-nearest element x of the f-induced fiber $F_{f(z)}$. Here

$$x^T = [x_1^T x_2^T \ldots x_N^T], \quad x_0^T = [x_{01}^T x_{02}^T \ldots x_{0N}^T] \quad (A1)$$

and in the most general case the three-dimensional transformation sought can be expressed pointwise by the *similarity transformation*

$$x_i = \lambda R(\theta) z_i + t = x_i(z_i, p), \quad p^T = [\theta^T t^T \lambda] \quad (A2)$$

where the transformation parameters p consist of three rotational parameters $\theta = [\theta_1\theta_2\theta_3]^T$ defining the orthogonal matrix $R = R(\theta)$, three parallel displacement parameters $t = [t_1 t_2 t_3]^T$, and a scale parameter λ.

The problem is to find the optimal values of θ, t, λ which minimize $f = f(p) = (x - x_0)^T(x - x_0)$. We first note that for

$$\dot{R} = \frac{\partial}{\partial a}R \quad .$$

$$RR^T = I \Rightarrow \dot{R}R^T + R\dot{R}^T = 0$$
$$\Rightarrow \dot{R}R^T = -(\dot{R}R^T)^T \equiv [\omega\times] \Rightarrow \dot{R} = [\omega\times]R$$

This can be applied for the derivatives with respect to θ_k by setting $\frac{\partial}{\partial\theta_k}R = [\omega_k\times]R$.

$$\frac{\partial x_i}{\partial\theta_k} = \lambda\frac{\partial}{\partial\theta_k}Rz_i = \lambda[\omega_k\times]Rz_i = -\lambda[(Rz_i)\times]\omega_k = -\lambda R[z_i\times]R^T\omega_k \quad (A3)$$

$$\frac{\partial x_i}{\partial\theta} = \left[\frac{\partial x_i}{\partial\theta_1}\frac{\partial x_i}{\partial\theta_2}\frac{\partial x_i}{\partial\theta_3}\right] = -\lambda R[z_i\times]R^T[\omega_1\omega_2\omega_3] \equiv -\lambda R[z_i\times]R^T\Omega \quad (A4)$$

where

$$\frac{\partial x_i}{\partial t} = I, \qquad \frac{\partial x_i}{\partial\lambda} = Rz_i \quad (A5)$$

To minimize f we simply set

$$\frac{\partial f}{\partial p} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial p} = 2(x - x_0)^T\left[\frac{\partial x}{\partial\theta}\frac{\partial x}{\partial t}\frac{\partial x}{\partial\lambda}\right] = 0 \quad (A6)$$

or explicitly

$$\sum_i (x_i - x_{0i})^T\frac{\partial x_i}{\partial\theta} = \sum_i(\lambda Rz_i + t - x_{0i})^T(-\lambda R[z_i\times]R^T\Omega) = 0 \quad (A7)$$

$$\sum_i (x_i - x_{0i})^T\frac{\partial x_i}{\partial t} = \sum_i(\lambda Rz_i + t - x_{0i})^T I = 0 \quad (A8)$$

$$\sum_i (x_i - x_{0i})^T\frac{\partial x_i}{\partial\lambda} = \sum_i(\lambda Rz_i + t - x_{0i})^T(Rz_i) = 0 \quad (A9)$$

which assuming that $\lambda \neq 0$ and $|\Omega| \neq 0$, can be rewritten as

$$\lambda\sum_i[z_i\times]z_i + \sum_i[z_i\times]R^T t = \sum_i[z_i\times]R^T x_{0i} \quad (A10)$$

$$\lambda R\sum_i z_i + \sum_i t = \sum_i x_{0i} \quad (A11)$$

$$\lambda\sum_i z_i^T z_i + \sum_i z_i^T R^T t = \sum_i z_i^T R^T x_{0i} \quad (A12)$$

Taking into account that $[z_i\times]z_i = 0$ and setting

$$\bar{z} \equiv \frac{1}{N}\sum_i z_i, \qquad \bar{x}_0 \equiv \frac{1}{N}\sum_i x_{0i} \quad (A13)$$

Eqs. (A10)–(A12) take the form

$$N[\bar{z}\times]R^T t = \sum_i[z_i\times]R^T x_{0i} \quad (A14)$$

$$\lambda R\bar{z} + t = \bar{x}_0 \quad (A15)$$

$$\lambda\sum_i z_i^T z_i + N\bar{z}^T R^T t = \sum_i z_i^T R^T x_{0i} \quad (A16)$$

Solving Eq. (A15) for t

$$t = \bar{x}_0 - \lambda R\bar{z} \quad (A17)$$

and replacing in Eq. (A16) gives

$$\lambda = \frac{\sum_i z_i^T R^T(x_{0i} - \bar{x}_0)}{\sum_i z_i^T z_i - N\bar{z}^T\bar{z}} \quad (A18)$$

which can be also written in the form

$$\lambda = \frac{\sum_i (x_{0i} - \bar{x}_0)^T R(z_i - \bar{z})}{\sum_i (z_i - \bar{z})^T(z_i - \bar{z})} \quad (A19)$$

To determine R (in fact θ) we replace t in Eq. (A14)

$$N[\bar{z}\times]R^T\bar{x}_0 - \lambda N[\bar{z}\times]\bar{z} = \sum_i[z_i\times]R^T x_{0i} \quad (A20)$$

which in view of $[\bar{z}\times]\bar{z} = 0$ can be written as

$$\frac{1}{N}\sum_i[z_i\times]R^T x_{0i} - [\bar{z}\times]R^T\bar{x}_0 = \frac{1}{N}\sum_i[x_{0i}\times]Rz_i - [\bar{x}_0\times]R\bar{z} = 0 \quad (A21)$$

or in the equivalent form

$$\frac{1}{N}\sum_i [(\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)\times]\boldsymbol{R}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) = \boldsymbol{0} \tag{A22}$$

Equation (A22) is a nonlinear equation which can be solved to obtain the values of the parameters $\boldsymbol{\theta}$ in any particular parametrization of the rotation matrix $\boldsymbol{R}(\boldsymbol{\theta})$. The obtained values should be next substituted in Eq. (A19) in order to determine the value of the scale parameter λ. The similarity transformation can be realized using these values, once $\boldsymbol{t}$ and λ are replaced from Eq. (A17) and (A19), respectively, into Eq. (A2) to obtain

$$\boldsymbol{x}_i = \overline{\boldsymbol{x}}_0 + \frac{\sum_i (\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)^{\mathrm{T}}\ \boldsymbol{R}(\boldsymbol{z}_i - \overline{\boldsymbol{z}})}{\sum_i (\boldsymbol{z}_i - \overline{\boldsymbol{z}})^{\mathrm{T}}\ (\boldsymbol{z}_i - \overline{\boldsymbol{z}})}\boldsymbol{R}(\boldsymbol{\theta})(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A23}$$

In the case of the *rigid transformation* $\boldsymbol{x}_i = \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{z}_i + \boldsymbol{t}$ we obtain only Eds. (A7), (A8) with $\lambda = 1$ and following the same procedure as before we arrive at the solution

$$\frac{1}{N}\sum_i [(\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)\times]\boldsymbol{R}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) = \boldsymbol{0} \tag{A24}$$

$$\boldsymbol{x}_i = \overline{\boldsymbol{x}}_0 + \boldsymbol{R}(\boldsymbol{\theta})(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A25}$$

The planar (two-dimensional) case is somewhat different since the 2×2 rotation matrix $\boldsymbol{R}(\vartheta)$ depends only on a single parameter ϑ. Setting

$$\boldsymbol{W} \equiv \frac{\partial}{\partial\vartheta}\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \dot{\boldsymbol{R}}\boldsymbol{R}^{\mathrm{T}} \tag{A26}$$

where $\boldsymbol{W}$ is a 2×2 antisymmetric matrix, the only difference from the three-dimensional similarity transformation case (apart from the fact that the vectors $\boldsymbol{x}_i, \boldsymbol{x}_{0i}, \boldsymbol{z}_i, \overline{\boldsymbol{x}}_0, \overline{\boldsymbol{z}}$ involved are now two-dimensional) is that Eq. (A7) is replaced by

$$\sum_i (\boldsymbol{x}_i - \boldsymbol{x}_{0i})^{\mathrm{T}}\ \frac{\partial \boldsymbol{x}_i}{\partial\vartheta} = \sum_i (\lambda\boldsymbol{R}\boldsymbol{z}_i + \boldsymbol{t} - \boldsymbol{x}_{0i})^{\mathrm{T}}\ (\lambda\boldsymbol{W}\boldsymbol{R}\boldsymbol{z}_i) = \boldsymbol{0} \tag{A27}$$

and the final solution for the *planar similarity transformation* takes the form

$$\frac{1}{N}\sum_i (\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)^{\mathrm{T}}\ \boldsymbol{W}\boldsymbol{R}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) = \boldsymbol{0} \tag{A28}$$

$$\lambda = \frac{\sum_i (\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)^{\mathrm{T}}\ \boldsymbol{R}(\boldsymbol{z}_i - \overline{\boldsymbol{z}})}{\sum_i (\boldsymbol{z}_i - \overline{\boldsymbol{z}})^{\mathrm{T}}\ (\boldsymbol{z}_i - \overline{\boldsymbol{z}})} \tag{A29}$$

$$\boldsymbol{x}_i = \overline{\boldsymbol{x}}_0 + \lambda\boldsymbol{R}(\vartheta)(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A30}$$

We can proceed further with the solution by choosing the usual parametrization

$$\boldsymbol{R}(\vartheta) = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \Rightarrow \boldsymbol{W} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{A31}$$

With this particular choice Eq. (A28) takes the form

$$a\cos\vartheta = b\sin\vartheta \tag{A32}$$

where

$$\begin{aligned} a &\equiv \frac{1}{N}\sum_i (\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \\ b &\equiv \frac{1}{N}\sum_i (\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}}_0)^{\mathrm{T}}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \end{aligned} \tag{A33}$$

There are two solutions: $\cos\vartheta = \pm\frac{6}{\sqrt{a^2+b^2}}$, $\sin\vartheta = \pm\frac{6}{\sqrt{a^2+b^2}}$ (both positive or both negative). Replacing in Eq. (A29) we obtain

$$\lambda = \frac{b\cos\vartheta + a\sin\vartheta}{s_z^2} = \pm\frac{\sqrt{a^2+b^2}}{s_z^2} \tag{A34}$$

where

$$s_z^2 \equiv \frac{1}{N}\sum_i (\boldsymbol{z}_i - \overline{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A35}$$

With the obtained values of $\sin\vartheta, \cos\vartheta$ and λ, the planar similarity transformation of Eq. (A30) becomes

$$\boldsymbol{x}_i = \overline{\boldsymbol{x}}_0 + \frac{1}{s_z^2}\begin{bmatrix} b & a \\ -a & b \end{bmatrix}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A36}$$

In the case of the *planar rigid transformation*, the solution is given by

$$\frac{1}{N}\sum_i (\boldsymbol{x}_{0i} - \overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) = \boldsymbol{0} \tag{A37}$$

$$\boldsymbol{x}_i = \overline{\boldsymbol{x}}_0 + \boldsymbol{R}(\vartheta)(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A38}$$

and for the above specific parametrization of $\boldsymbol{R}(\vartheta)$ we obtain two solutions

$$\boldsymbol{x}_i = \overline{\boldsymbol{x}}_0 \pm \frac{1}{\sqrt{a^2+b^2}}\begin{bmatrix} b & a \\ -a & b \end{bmatrix}(\boldsymbol{z}_i - \overline{\boldsymbol{z}}) \tag{A39}$$

which correspond to two rotation angles ϑ and $\vartheta + \pi$. A further investigation is needed in each particular application, in order to determine which of the two solutions minimizes in fact the target function $f = (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_0)$.

Appendix B

Proof of the equivalence of the representations of Eqs. (81) and (82)

Our starting point is Eq. (81)

$$\begin{aligned} \boldsymbol{G}_0 = {} & \boldsymbol{S}_0[\boldsymbol{\Lambda}\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{S}_0]^{-1}[\boldsymbol{V}_1^{\mathrm{T}} - \boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{C}_0(\boldsymbol{V}_2^{\mathrm{T}}\boldsymbol{C}_0)^{-1}\boldsymbol{V}_2^{\mathrm{T}}] \\ & + \boldsymbol{D}_0[\boldsymbol{V}_2^T\boldsymbol{C}_0]^{-1}\boldsymbol{V}_2^T \end{aligned} \tag{B1}$$

We need an interpretation of this formula which will be independent of U_1, U_2, V_1, V_2 and depends only on S_0, C_0, D_0, and related matrices.

Let us consider in addition to the matrix C with column span $R(C)$ a matrix $C^\perp$ with column span $R(C)^\perp = R(C)^\perp$. This means that $C^\perp$ must be such that

$$\text{rank}[C|C^\perp] = n \quad \text{and} \quad C^T C^\perp = 0$$

Lemma. A possible choice for $C^\perp$ is

$$C^\perp = \begin{bmatrix} I \\ -C_2^{-T} C_1^T \end{bmatrix} \tag{B2}$$

Proof. Setting $C^\perp = \begin{bmatrix} A \\ B \end{bmatrix}$ we have

$$C^T C^\perp = [C_1^T, C_2^T] \begin{bmatrix} A \\ B \end{bmatrix} = C_1^T A + C_2^T B = 0$$

yielding $B = -C_2^{-T} C_1^T A$ while A must be chosen so that

$$\text{rank}[C|C^\perp] = \text{rank}\begin{bmatrix} C_1 & A \\ C_2 & -C_2^{-T} C_1^T A \end{bmatrix} = \text{rank}\begin{bmatrix} A & C_1 \\ -C_2^{-T} C_1^T A & C_2 \end{bmatrix} = n$$

With the obvious simplest choice $A = I_r$ we obtain

$$C^\perp = \begin{bmatrix} I \\ -C_2^{-T} C_1^T \end{bmatrix}. \qquad \square$$

Returning to the original bases we have $0 = C^T C^\perp = C_0^T V C^\perp$, implying that

$$C_0^\perp = VC^\perp = V_1 - V_2 C_2^{-T} C_1^T = V_1 - V_2 (V_2 C_0)^{-T} C_0^T V_1$$

The range $R(f)$ is represented by $R(F) = R(FS)$ since $f(S) = R(f)$ and

$$FS = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} \Lambda S_1 \\ 0 \end{bmatrix}$$

Lemma. A possible choice for $(FS)^\perp$ is

$$(FS)^\perp = \begin{bmatrix} 0 \\ I_d \end{bmatrix} \tag{B3}$$

Proof. Setting $(FS)^\perp = \begin{bmatrix} A \\ B \end{bmatrix}$ for any matrix with column-span $R(FS)^\perp$ we have

$$0 = (FS)^T\ (FS)^\perp = [S_1^T \Lambda\ 0] \begin{bmatrix} A \\ B \end{bmatrix} = S_1^T \Lambda\ A = 0$$

which yields $A = 0$, while B must be chosen so that

$$\text{rank}\left[(FS)|(FS)^\perp\right] = \text{rank}\begin{bmatrix} \Lambda S_1 & 0 \\ A & B \end{bmatrix} = \text{rank}\begin{bmatrix} \Lambda S_1 & 0 \\ A & B \end{bmatrix} = m.$$

With the obvious simplest choice $B = I_d$ we obtain

$$(FS)^\perp = \begin{bmatrix} 0 \\ I_d \end{bmatrix}. \qquad \square$$

Returning to the original bases we have

$$0 = (FS)^T\ (FS)^\perp = (V^T F_0 U U^T S_0)^T\ (FS)^\perp = (F_0 S_0)^T\ V(FS)^\perp$$

implying that a possible choice for $(F_0 S_0)^\perp$ is

$$(F_0 S_0)^\perp = V(FS)^\perp = [V_1 V_2] \begin{bmatrix} 0 \\ I_d \end{bmatrix} = V_2$$

Using the derived relations

$$C_0^\perp = V_1 - V_2 (V_2 C_0)^{-T} C_0^T V_1 \tag{B4}$$

$$(F_0 S_0)^\perp = V_2 \tag{B5}$$

and

$$(F_0 S_0) = VFU^T US = V(FS) \tag{B6}$$

we obtain

$$\begin{aligned}(C_0^\perp)^T F_0 S_0 &= [V_1^T - V_1^T C_0 (V_2^T C_0)^{-1} V_2^T][V_1 V_2] \begin{bmatrix} \Lambda S_1 \\ 0 \end{bmatrix} \\ &= [V_1^T - V_1^T C_0 (V_2^T C_0)^{-1} V_2^T] V_1 \Lambda S_1 \\ &= V_1^T V_1 \Lambda S_1 \\ &\quad - V_1^T C_0 (V_2^T C_0)^{-1} V_2^T V_1 \Lambda S_1 \\ &= \Lambda S_1 = \Lambda U_1^T S_0 \end{aligned} \tag{B7}$$

where we have used the fact that $V_1^T V_1 = I$ and $V_2^T V_1 = 0$ as a consequence of $V^T V = I$. Utilizing the derived relation $\Lambda U_1^T S_0 = (C_0^\perp)^T\ F_0 S_0$ and the one for $(C_0^\perp)^T$ in Eq. (B4) the generalized inverse g is represented in the original bases by the matrix

$$\begin{aligned} G_0 = &\ S_0[(C_0^\perp)^T\ F_0 S_0]^{-1} (C_0^\perp)^T \\ &+ D_0[((F_0 S_0)^\perp)^T\ C_0]^{-1} ((F_0 S_0)^\perp)^T \end{aligned} \tag{B8}$$

This equation is identical with the representation of Eq. (82) derived by Teunissen (1985, Eq. 2.7).

References

Baarda W (1973) S-Transformations and criterion matrices. Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 5, No. 1

Bjerhammar A (1951) Rectangular reciprocal matrices with special references to geodetic calculations. Bull Geod 52: 188–220

Blaha G (1971) Inner adjustment constraints with emphasis on range observations. Dept Geodetic Science Report No. 148, The Ohio State University

Choquet-Bruhat Y, DeWitt-Morette C, Dillard-Bleick M (1977) Analysis manifolds and physics. North-Holland, Amsterdam

Dermanis A (1991) A unified approach to linear estimation and prediction. Presented at the IUGG XXth General Assembly,

International Association of Geodesy, Vienna, August 11–24, 1991
Dermanis A, Grafarend E (1981) Estimability analysis of geodetic, astrometric and geodynamical quantities in Very Long Baseline Interferometry. Geophys J R Astronom Soc 64: 31–56
Dermanis A, Sanso F (1995) Nonlinear estimation problems for nonlinear models. Manuscr Geod 20: 110–122
Grafarend E (1974) Optimization of geodetic networks. Boll Geod Sci Affini 33: 351–406
Grafarend E, Kampmann G (1996) $C_{10}(3)$: the ten-parameter conformal group as a datum transformation in three-dimensional Euclidean space. Z Vermessungswesen 121: 68–77
Grafarend E, Schaffrin B (1974) Unbiased free net adjustment. Surv Rev XXII, 171: 200–218
Grafarend E, Schaffrin B (1976) Equivalence of estimable quantities and invariants in geodetic networks. Z Vermessungswesen 101: 485–491
Grafarend E, Schaffrin B (1991) The planar trisection problem and the impact of curvature on non-linear least squares estimation. Comput Stat Data Anal 12: 187–199
Halmos PR (1974) Finite-dimensional vector spaces. Springer, Berlin Heidelberg New York
Koch K-R (1982) S-transformations and projections for obtaining estimable parameters. In: Blotwijk, M.J., F.J.J. Brouwer, D.T. van Daalen, H.M. de Heus, J.T. Gravesteijn, H.C. van der Hoek, J.J. Kok, P.J.G. Teunissen (eds) 40 Years of Thought, Anniversary volume for Prof. Baarda's 65th Birthday Vol. 1, Technische Hogeschool Delft, Delft, pp 136–144
Lohse P (1994) Ausgleichungsrechnung in nichtlinearen modellen. Deutsche Geod Komm C 429
Loomis LH, Sternberg S (1980) Advanced calculus. Addison-Wesley, Reading Mass
Meissl P (1965) Über die Innere Genauigkeit dreidimensionaler Punkthaufen. Z Vermessungswesen 90: 109–118
Meissl P (1969) Zusammengfassung und Ausbau der inneren Fehlertheorie eines Punkthaufens. Deutsche Geod Komm A 61: 8–21
Mierlo, J van (1980) Free network adjustment and S-transformations. Deutsche Geod Komm B 252: 41–54
Moore EH (1920) On the reciprocal of the general algebraic matrix (abstract). Bull Amer Math Soc 26: 394–395
Penrose R (1955): A generalized inverse for matrices. Proc Cambridge Philos Soc 52: 17–19
Rao CR, Mitra SK (1971) Generalized inverse of matrices and its applications. Wiley, New York
Schaffrin B, Heidenreich E, Grafarend E (1977) A representation of the standard gravity field. Manuscr Geod 2: 135–174
Sternberg S (1964) Lectures on differential geometry. Prentice-Hall, Englewood Cliffs, N.J.
Takos I (1976) Generalized inverse of matrices and its applications (in Greek). Bull Hellenic Army Geogr Service 1976: 77–133
Teunissen PJG (1985) The geometry of geodetic inverse linear mapping and non-linear adjustment. Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 8, No. 1
Teunissen PJG (1989a) Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity. In: Brunner FK, Rizos C (eds) Developments in four-dimensional geodesy. Lecture Notes in Earth Sciences, No. 29. Springer, Berlin Heidelberg New York pp 241–264
Teunissen PJG (1989b) First and second moments of non-linear least-squares estimators. Bull Geod 63: 253–262
Teunissen PJG (1990) Non-linear least squares. Manuscr Geod 15: 137–150
Xu P (1995) Testing the hypotheses of non-estimable functions in free net adjustment models. Manuscr Geod 20: 73–81