Navigation in a model space with misfit-induced curvature

Kris Innanen

ABSTRACT

We investigate the consequences in inversion of adapting our geometrical picture of model space. We consider that model space is curved by the presence of the misfit or objective function. Within such a space, the simplest possible curves, equivalent to straight lines in flat space, are geodesics. Paths are normally followed by model estimates in the process of updating and in estimating uncertainty, and we examine the geodesics with this in mind — is it possible that visiting points in model space in the order determined by a geodesic in these new spaces is a valuable exercise? There are some hints that the answer may be yes. Geodesics computed with coordinates with lower (covariant) indices appear to orbit regions of model space surrounding the minimum of the objective function in a manner which may have application in uncertainty evaluation.

INTRODUCTION

The possibility of analyzing elements of model space by using more a general Riemannian geometry than is standard in geophysical inverse problems was introduced in the 2020 CREWES report. There, flat but non-Cartesian (i.e., oblique-rectilinear) coordinate systems were focused on, but the use of curvature as a way of bringing the misfit function into the problem was introduced. In this paper that possibility is fleshed out and its basic numerical behaviour in low dimensions is examined.

Typically in a geophysical inverse problem we envision a flat model space, overlain by a Cartesian coordinate system, in which individual models take the form of position vectors. One vector is selected as a starting point, and then other update vectors are added to the starting vector in search of an optimal final vector. This forms a discrete path through model space. The updates are decided by the derivatives of a scalar functional defined on model space called the objective function. Here we suppose instead that model space is no longer flat, but is in fact curved, and the curvature is provided by the objective function. After the space has been properly curved, we no longer explicitly analyze it, but instead ask about what simple paths in that newly-curved space behave like. The main question is, are these paths useful, in the sense of model construction or uncertainty characterization?

In this paper, we set the problem up mathematically, and carry out some initial numerical exploration of the idea. Therefore we do not conclude with new algorithms for inversion; rather, a set of observations about the character of simple paths (geodesics) in this space. The formalism of geodesics and curvature is reviewed in a companion paper in this 2021 report.

Notation

Tensors are labelled with either Roman or Greek indices; upper indices refer to the contravariant components of a vector and lower indices refer to the covariant components. The number of unrepeated indices equals the rank of the tensor. Repeated indices (one up and one down) imply summation, e.g., $a^\mu a_\mu = a^1 a_1 + ... a^N a_N$. Tensor indices are raised and lowered in general with the metric tensor and its inverse $m_{\mu\nu}$ and $m^{\mu\nu} = (m_{\mu\nu})^{-1}$, for instance $b^\mu = m^{\mu\nu} b_\nu$ and $c_\mu = m_{\mu\nu} c^\nu$.

MODEL SPACE WITH CURVATURE INDUCED BY A QUADRATIC OBJECTIVE FUNCTION

Consider the N dimensional model space s and a parameterization, or coordinate system

$$s^\mu = \begin{bmatrix} s^1 \\ s^2 \\ \vdots \\ s^N \end{bmatrix}. \tag{1}$$

Further, consider a quadratic objective function

$$\Phi(s) = \frac{1}{2} s^\mu \phi_{\mu\nu} s^\nu - \varphi_\lambda s^\lambda + \psi, \tag{2}$$

whose minimum is the solution of an inverse problem of interest. In (2) the matrix $\phi_{\mu\nu}$ is referred to as the Hessian. The minimum is often determined iteratively through model updates based on the gradient direction

$$g_\mu = \Phi_{,\mu} = \phi_{\mu\nu} s^\nu - \varphi_\mu. \tag{3}$$

For later convenience we define g as the length of the gradient vector, and h and h_μ as

$$h = g^\alpha \phi_{\alpha\beta} g^\beta, \text{ and } h_\mu = g^\nu \phi_{\mu\nu}. \tag{4}$$

We will require the metric tensor to be in-hand in order to compute the quantities in (4), as they require known quantities to have their indices raised or lowered.

Typically, we treat Φ as an external object whose derivatives are suggestive about directions in model space along which to update. Here we will instead have Φ induce in model space an intrinsic, geometric curvature, and examine how simple paths in such a curved s behave.

The quadratic objective function is a good representative and tractable example of a figure of merit or target for optimization. This is because even in nonlinear cases the objective function in the vicinity of the minimum is often approximately quadratic, and furthermore many different kinds of optimization produce variants of the same form (i.e., equation 2). We could let $\Phi(s)$ contain the full objective information i.e., be made up of both data misfit and regularization terms:

$$\Phi(s) = \Phi_d(s) + \beta \Phi_m(s), \tag{5}$$

where, for instance, if in the forward modelling scheme the ith datum is $d^i = F^i_\mu s^\mu$,

$$\Phi_d(s) = \frac{1}{2} s^\mu \big(F^i_\mu F_{i\nu}\big) s^\nu - \big(F^i_\lambda d_i\big) s^\lambda + \frac{1}{2} d^i d_i, \tag{6}$$

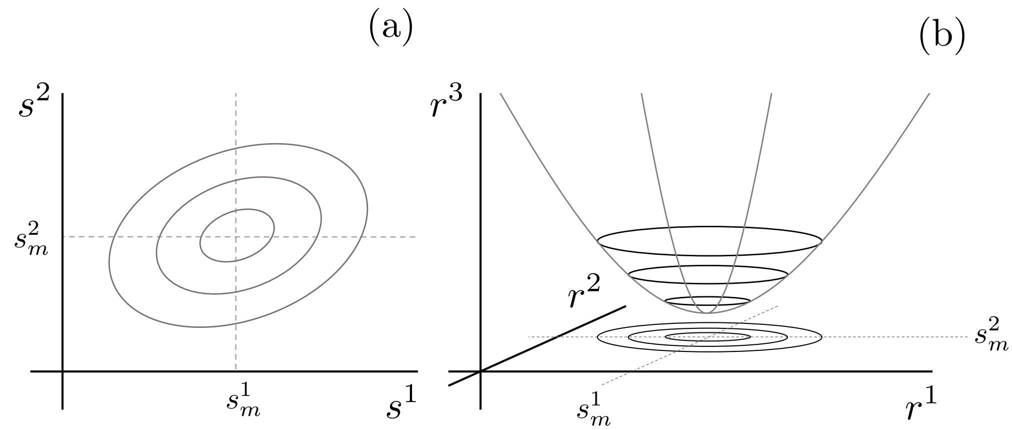

FIG. 1. Model space curved by misfit, N=2 dimensional example. (a) Standard view of model space s as a plane overlain by a Cartesian coordinate system (contours of a quadratic objective function included). (b) The same 2D model space viewed as a paraboloid surface, curved by the objective function. The curvature is quantified and visualized by introducing a flat $N + 1 = 3$ dimensional space r, within which the curved model space is embedded. The higher dimensional space r can be discarded after the metric tensor is determined.

and if we penalize model properties magnified by an operator $W_{\mu\nu}$,

$$\Phi_m(s) = \frac{1}{2} s^\mu \big(W_\mu^\lambda W_{\nu\lambda}\big) s^\nu. \tag{7}$$

Or, we could set $\Phi = \Phi_m$, i.e., to be only the contribution of the regularization term to the objective function. In the latter case, “straight line” navigation through model space would be affected only by prior information, rather than by a data set. The choice impacts the detailed form of the g and h quantities in (3)-(4), and so the general developments to follow are not sensitive to it.

Metric of a model space curved by an objective function

We introduce an auxiliary (flat) space r, of $N + 1$ dimensions, with Cartesian coordinates

$$r^n = \begin{bmatrix} r^1 \\ r^2 \\ \vdots \\ r^N \\ r^{N+1} \end{bmatrix}. \tag{8}$$

Let s be a surface embedded in r (Figure 1) satisfying the equations

$$r^1 = s^1, \;\; r^2 = s^2, \quad \ldots, \;\; r^N = s^N, \;\; r^N = \phi. \tag{9}$$

This produces in s the intrinsic curvature discussed abo e. To determine the metric tensor $m_{\mu\nu}(s)$ of this curved s, we observe that contravariant components of vectors lying in both

r and s transform as $r^n = r^n_{,\nu} s^\nu$, where $r^n_{,\nu}$ is the $N \times (N+1)$ matrix

$$r^n_{,\nu} = \frac{\partial r^\mu}{\partial s^\nu} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \\ g_1 & g_2 & g_3 & \dots & g_N \end{bmatrix}. \tag{10}$$

From this we have

$$m_{\mu\nu} = r_{n,\mu} r^n_{,\nu} = \begin{bmatrix} 1+g_1^2 & g_1 g_2 & \dots & g_1 g_N \\ g_2 g_1 & 1+g_2^2 & \dots & g_2 g_N \\ & \ddots & & \\ g_N g_1 & g_N g_2 & \dots & 1+g_N^2 \end{bmatrix}. \tag{11}$$

The metric tensor can also be inverted using the identity $(\mathbf{I}+\mathbf{u}\mathbf{u}^T)^{-1} = \mathbf{I} - \mathbf{u}\mathbf{u}^T/(1+\mathbf{u}^T\mathbf{u})$ to produce its index-raising form:

$$m^{\mu\nu} = \begin{bmatrix} 1-\frac{g_1^2}{1+g^2} & -\frac{g_1 g_2}{1+g^2} & \dots & -\frac{g_1 g_N}{1+g^2} \\ -\frac{g_2 g_1}{1+g^2} & 1-\frac{g_2^2}{1+g^2} & \dots & -\frac{g_2 g_N}{1+g^2} \\ & \ddots & & \\ -\frac{g_N g_1}{1+g^2} & -\frac{g_N g_2}{1+g^2} & \dots & 1-\frac{g_N^2}{1+g^2} \end{bmatrix}, \tag{12}$$

where $g^2 = g^\mu g_\mu$. After summing over n to create the right-hand side of (11), no dependency on r remains, and per standard procedure the higher-dimensional space can now be discarded, and $m_{\mu\nu}(s)$ be considered a received quantity. In other words, we now view ourselves as inhabitants of a space s within which lengths must be measured using $m_{\mu\nu}$. Since the components of the gradient depend on position in model space (from 3), the metric tensors do also, so, to us as inhabitants of s, the space appears curved. At present we cannot yet be certain whether this curvature is apparent or intrinsic to the space.

Curvature

With the spatially-varying metric tensor in hand, Christoffel symbols of the first and second kind for this curved model space can also be directly computed. By construction we observe that they take the form of Kronecker products of the gradient and the Hessian:

$$\Gamma_{\mu\nu\lambda} = g_\lambda \phi_{\mu\nu}, \text{ and } \Gamma^\lambda_{\mu\nu} = g^\lambda \phi_{\mu\nu}. \tag{13}$$

The curvature and Ricci tensors and the scalar curvature are likewise determined directly in terms of the Christoffel symbols. The curvature tensor is

$$R^\beta_{\nu\rho\sigma} = \phi^\beta_\rho \phi_{\nu\sigma} - \phi^\beta_\sigma \phi_{\nu\rho} + \Gamma^\beta_{\nu\sigma} h_\rho - \Gamma^\beta_{\nu\rho} h_\sigma, \tag{14}$$

the Ricci tensor and scalar curvature are the contractions

$$R^\nu_\rho = \phi^\mu_\rho \phi^\lambda_\mu - \phi^\mu_\mu \phi^\lambda_\rho + h_\nu h_\rho - h\phi_{\nu\rho}, \; R = \phi^\rho_\beta \phi^\beta_\rho - \phi^\beta_\beta \phi^\rho_\rho + h^\rho h_\rho - h\phi^\rho_\rho. \tag{15}$$

Geodesics

A geodesic is a position vector varying continuously with a single scalar parameter, normally the path length τ, i.e., $s^{\mu}(\tau)$, satisfying

$$\frac{d^2 s^{\lambda}}{d\tau^2} + \Gamma^{\lambda}_{\mu\nu}\frac{ds^{\mu}}{d\tau}\frac{ds^{\nu}}{d\tau} = 0, \tag{16}$$

or

$$\frac{du^{\lambda}}{d\tau} + \Gamma^{\lambda}_{\mu\nu}u^{\mu}u^{\nu} = 0, \tag{17}$$

with $u^{\mu} = ds^{\mu}/d\tau$ being the tangent vector to the path. Geodesic in the model space we have been describing therefore satisfy (substituting 13)

$$\frac{du^{\lambda}}{d\tau} + g^{\lambda}\phi_{\mu\nu}u^{\mu}u^{\nu} = 0. \tag{18}$$

From this a simple first-order discrete solution for the ith position and tangent vectors along the geodesic, given an initial position $s^{\lambda}(0)$ and take-off direction $u^{\lambda}(0)$, is obtained:

$$\begin{aligned} u^{\lambda}(i+1) &= u^{\lambda}(i) - \Delta\tau g^{\lambda}(i)u^{\mu}(i)\phi_{\mu\nu}(i)u^{\nu}(i), \\ s^{\lambda}(i+1) &= s^{\lambda}(i) + \Delta u^{\lambda}(i), \end{aligned} \tag{19}$$

where $\Delta\tau$ is the discrete step length along the geodesic.

The geodesic equation with lowered index

We can make a version of the equation in which the particle track is computed in terms of lowered indices:

$$m_{\lambda\sigma}\left(\frac{du^{\sigma}}{d\tau} + g^{\sigma}\phi_{\mu\nu}u^{\mu}u^{\nu}\right) = 0, \tag{20}$$

or

$$\frac{du_{\lambda}}{d\tau} + g_{\lambda}\phi_{\mu\nu}u^{\mu}u^{\nu} = 0, \tag{21}$$

and its discretized form

$$\begin{aligned} u_{\lambda}(i+1) &= u_{\lambda}(i) - \Delta\tau g_{\lambda}(i)u^{\mu}(i)\phi_{\mu\nu}(i)u^{\nu}(i), \\ s_{\lambda}(i+1) &= s_{\lambda}(i) + \Delta u_{\lambda}(i). \end{aligned} \tag{22}$$

The coordinate pairs produced by this equation are numerically different from those produced by the geodesic, so to avoid confusion we will refer to this track as the misfit-orbit. The misfit-orbit appears to have properties that are more interesting from an inverse problem point of view than those of the standard geodesic, so we will focus on its behaviour in the examples below.

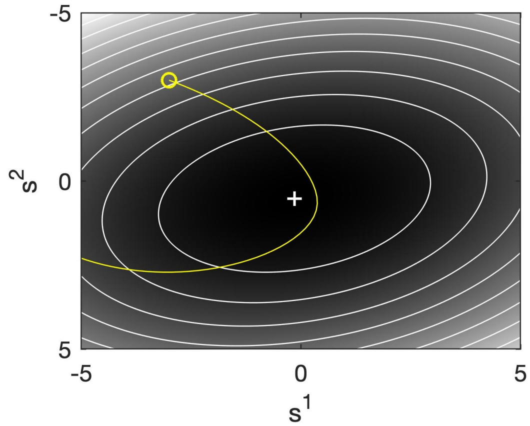

FIG. 2. A geodesic computed via equation (19).

LOW DIMENSIONAL BEHAVIOUR OF A MISFIT-ORBIT

Here we examine some of the numerical behaviour of moving points in model space. We will focus on the coordinate pairs generated by the geodesic equation with lowered indices, which are related to geodesics.

A bivariate example

We will remain largely in a bivariate environment in this paper, in order to be able to make easily-visualizable results. Thus, the objective function will be explicitly built with the matrix and vector pair

$$\phi_{\mu\nu} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix}, \quad \varphi_{\mu} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix}, \tag{23}$$

as well as a scalar:

$$\Phi(s) = \frac{1}{2}[s^1 \; s^2] \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix} \begin{bmatrix} s^1 \\ s^2 \end{bmatrix} - [s^1 \; s^2] \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \psi. \tag{24}$$

So, a particular case can be produced by choosing 6 numbers.

The upper-index position vector components of the orbit

Let us start with the geodesic equation. Tracks which obey this equation should be expected to be straight lines that are perturbed into smooth curves by the space. This is confirmed with simple numerical implementations of equation (19) with the 6 weights in equation (23). For example in Figure 2 a track is begun in a direction to the right of the minimum of the objective function, but it gradually deviates towards the minimum, prior to being ejected.

This clearly demonstrates that a geodesic is a path which, if followed by a model point as dictated by a particular inverse problem, gives us information about the minimum, and

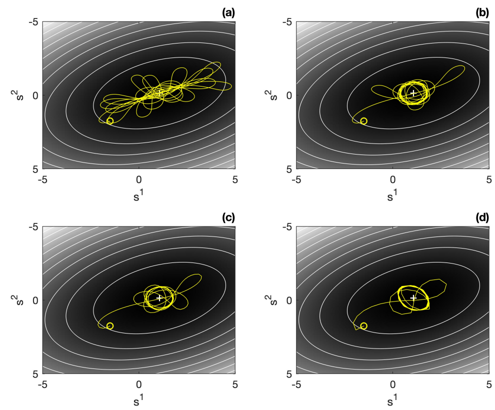

FIG. 3. Change in orbit as computed with the geodesic equation with index lowered; from (a) to (d) the discretization of the path grows.

by variation of the path, even uncertainty. However, the numerical values of the coordinates plotted in Figure 2 and their relationship to the inverse problem appear (qualitatively) complex and difficult to interpret. Plots of the numerical values of the coordinate pairs produced by solving equation (22), i.e., with lowered-indices, appear to produce more readily-interpretable tracks, so we move to this equation next.

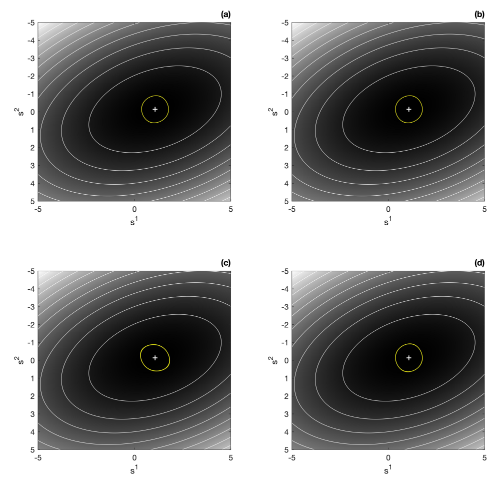

FIG. 4. Change in orbit as computed with the geodesic equation with index lowered; from (a) to (d) the discretization of the path grows. In this case only the second half of the full orbit length is included. We observe that the orbit has settled into a very regular, and almost precisely circular, shape.

The lower-index position vector components of the orbit

Misfit orbits satisfying equation (22) appear to be more locally constrained by the minimum. Notice that the coordinates of the minimum do not change when we move from plots with upper indices to ones with lower indices, because at that point the metric tensor lapses to the identity.

In fact, we observe behaviour that suggestive of an attractor, but one whose features are most strongly dictated by the discretization parameter $\Delta\tau$. In Figure 3, we show an example of this behaviour. The yellow circle is the starting model, and the misfit-orbit is begun with a take-off direction away from the minimum. The orbit is then computed with four different values of $\Delta\tau$, growing from small to large (a-d).

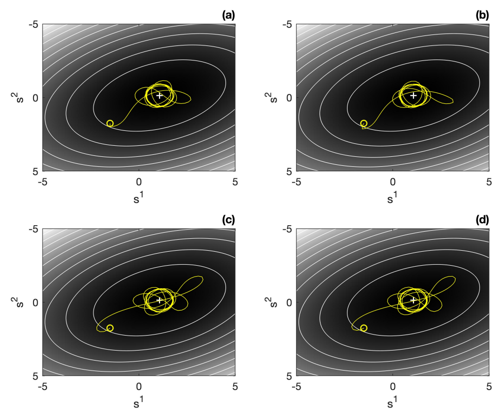

FIG. 5. Change in orbit as computed with the geodesic equation with index lowered; from (a) to (d) the take-off angle of the path is varied.

Very fine discretizations – i.e., very accurate estimates of the orbit – reveal complex behaviour in the vicinity of the minimum in the early history of the orbit, caused by the eliipticity of the objective function and its overall attractive influence. Examination of Figure 3a-d furthermore make clear how dependent the details of those paths are on However, although it is evident that the path in the initial steps away from the starting point appears to be strongly influenced by the discretization, after a certain number of iterations a surprisingly regular orbit is attained, and stuck to. To expose this, in Figures 4a-d the examples with varying path discretization are again plotted, but this time with the first half of the positions calculated not included. In all cases, by halfway through their full extent, the orbits have converged to very tight circles, with little inter-variation.

Dependence on take-off direction

This convergence does not appear to be particularly sensitive to take-off direction. In Figure 5a-d the take-off direction is varied, with all else kept fixed, and we see that the same final orbit is arrived at.

Dependence on initial position

The initial position does not appear to lead to strong variations in the final orbit either (Figure 6).

CONCLUSIONS

The possibilities associated with a treatment of model space as being curved by a misfit or objective function of interest are pursued in an initial way. The tracks produced by a the geodesic equation with indices lowered appear to produce some interesting behaviour, most of which is yet to be completely analyzed. The similarity between the objective function and an attractive potential was initially suggestive that orbits that probe regions of model

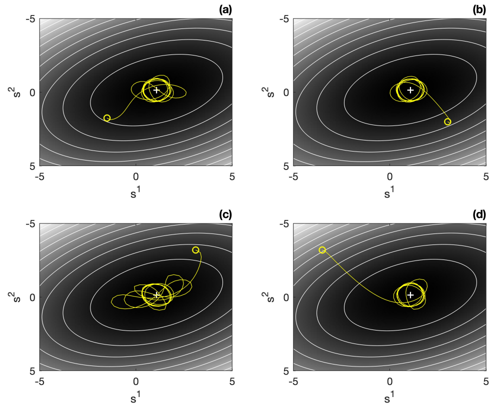

FIG. 6. Change in orbit as computed with the geodesic equation with index lowered; from (a) to (d) the take-off angle of the path is varied.

space, rather than say point to a single model, were likely to emerge. That appears to be the case, meaning that at present the treatment seems to be most immediately applicable to uncertainty characterization. The geodesic equation is nonlinear, and so analytic interpretation of the orbits can be difficult. The tendency of the coordinates of the solution to the geodesic equation with lowered indices appears to naturally fall into a non-chaotic, qualitatively circular, attractor in a bivariate setting is notable. Whether this persists at higher dimensions, and whether we can make practical use of it, is a matter of ongoing study. The use of an orbit like this as a suite of null-space shuttles is one possibility.

ACKNOWLEDGEMENTS

The sponsors of CREWES are gratefully thanked for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19, and in part by an NSERC-DG.