Model re-parameterization via misfit-based coordinate transforms

Kris Innanen

ABSTRACT

Re-parameterizations of model space, which are widely applied in seismic inverse problems, are coordinate transforms between non-Cartesian systems. To develop this we identify objective functions as scalar functions of the model vectors, which themselves are contravariant vectors; gradients of the objective function are covariant vectors. A procedure for general transformation to a pre-defined coordinate system and optimization within that system is set out. We argue that a under a class of transformations constrained by the Hessian operator in the reference system, steepest-descent updates are precisely parallel to Gauss-Newton updates, and, provided the transform can be efficiently determined, optimization within the transformed system should have favourable convergence properties. This class of transforms includes an infinite number of variants, and seeking examples from within this class with other, additional, favourable features appears warranted.

INTRODUCTION

In this year's CREWES report there are a set of papers on the general subject of design of transformations between coordinate systems, with application to increased optimization efficiency in inverse problems. This paper is meant to act as the central document, with the others being offshoots. The first paper (Innanen, 2020a) is a review of and/or introduction to some of the tools of tensor mathematics that geophysicists are less likely to have encountered, especially relating to the separate treatment of contravariant and covariant vector and tensor components. This is the second paper, where the basic ideas are set out. The third and fourth papers (Innanen, 2020b,c) are applications of the idea to two somewhat distinct kinds of problem: a low-dimensional problem, AVO inversion, and a high-dimensional problem, FWI. The fifth paper (Innanen, 2020d) provides detailed descriptions of numerical procedures for determining general transformation matrices based on the ideas in the other papers.

Because so much detail is shunted to other documents, this paper is quite short. In the first section, several features of quadratic objective functions as scalar functions of contravariant model vectors, gradients of these functions, and updates based on these gradients, are set out. In the second section, the conclusions of the first section motivate the formulation of a design scheme for coordinate systems in which steepest descent updates are as efficient as Newton updates. Finally, anticipating that these procedures will be particularly useful for large scale problems, several numerical features of the transform design problem are discussed.

OPTIMIZATION IN GENERAL RECTILINEAR COORDINATE SYSTEMS

Quadratic objective functions and their transformation

Let t^{μ}_{ν} be a transformation between rectilinear coordinate systems s and r, such that position vectors transform as

$$s^{\mu} = t^{\mu}_{\nu} r^{\nu}. \tag{1}$$

Let Φ be a scalar quadratic objective function

$$\Phi = s^{\mu} \phi_{\mu\nu} s^{\nu} - s^{\mu} \varphi_{\mu} + \gamma. \tag{2}$$

Substituting (1) into (2), we observe that under transformation to the r system Φ becomes

$$\Phi' = r^{\mu} \phi'_{\mu\nu} r^{\nu} - r^{\mu} \varphi'_{\mu} + \gamma', \tag{3}$$

where $\phi'_{\mu\nu} = t^{\lambda}_{\mu}\ \phi_{\lambda\sigma}\ t^{\sigma}_{\nu}$, $\varphi' = t^{\lambda}_{\mu}\ \varphi_{\lambda}$, and $\gamma' = \gamma$. The objective function Φ is a scalar, meaning that if s^{μ} and r^{μ} satisfy equation (1),

$$\Phi(s^{\mu}) = \Phi'(r^{\mu}). \tag{4}$$

Consider a fixed point r^{μ}_{0} and its associated $\Phi_0 = \Phi'(r^{\mu}_{0})$. By varying t^{μ}_{ν}, we can send this r^{μ}_{0} over into a range of output vectors $s^{\mu} = t^{\mu}_{\nu} r^{\nu}_{0}$. However, because of (4), this range is not unlimited, but is restricted to s^{μ} vectors for which $\Phi(s^{\mu}) = \Phi_0$. If we add to this that there exists only one s^{μ} whose associated Φ_0 is the minimum, it follows that if r^{μ}_{min} is the minimizer of Φ in the r system, $s^{\mu}_{\text{min}} = t^{\mu}_{\nu} r^{\nu}_{\text{min}}$ is the minimizer of Φ in the s system, independent of t^{μ}_{ν}. This means that if we find the minimizer in any coordinate system, we have it in all coordinate systems, independent of t^{μ}_{ν}.

Gradient vectors and their transformation

The gradient of the objective function in equation (2) is

$$\Phi_{,\mu}(s) = \frac{\partial \Phi(s)}{\partial s^{\mu}} = 2\phi_{\mu\nu} s^{\nu} - \varphi_{\mu}. \tag{5}$$

If a coordinate transform $s^{\mu} = t^{\mu}_{\nu} r^{\nu}$ is set up, a gradient, whose components are contravariant, transforms as

$$\Phi_{,\mu}(r) = t^{\nu}_{\mu} \Phi_{,\nu}(r) = 2 t^{\nu}_{\mu} \phi_{\nu\lambda} s^{\lambda} - t^{\sigma}_{\mu} \varphi_{\sigma}. \tag{6}$$

A step in the direction of the negative of the gradient in the s system is

$$\Delta s^{\mu} = -g^{\mu\nu} \Phi_{,\nu} = -g^{\mu\nu} \left[2\phi_{\nu\lambda} s^{\lambda} - \varphi_{\nu} \right] = \varphi^{\mu} - 2\phi^{\mu}_{\nu} s^{\nu}, \tag{7}$$

where $g^{\mu\nu}$ is the metric tensor in the s system. Since Φ is precisely quadratic, we can compare this against the displacement from any point s^{μ} to the minimum, which is

$$s^{\mu}_{\text{min}} - s^{\mu} = -\left(\phi^{-1}\right)^{\mu\nu} \Phi_{,\nu} = \left(\phi^{-1}\right)^{\mu\nu} \varphi_{\nu} - 2 s^{\mu}, \tag{8}$$

where $(\phi^{-1})^{\mu\lambda} \phi_{\lambda\nu} = \delta^{\mu}_{\nu}$. Comparing (7) and (8) it is clear that the steepest descent direction only points towards the minimum if $\phi_{\mu\nu} = \delta_{\mu\nu}$.

Transformation of steepest-descent updates

The shape and character of quadratic objective functions, and the directions of gradients from fixed starting points, can be changed with transformations of the type in (1). A standard steepest-descent procedure is as follows. We choose a starting point s_0^μ and at that point compute the gradient:

$$\Phi_{,\mu}(s_0) = \left.\frac{\partial \Phi}{\partial s^\mu}\right|_{s_0}. \tag{9}$$

We then scale the negative of the gradient with a coefficient α, usually determined through a line search, and update the starting point:

$$s_1^\mu = s_0^\mu + \Delta s^\mu = s_0^\mu - \alpha\, g^{\mu\lambda}\Phi_{,\lambda}(s_0). \tag{10}$$

The association of the gradient direction directly with a step, or displacement, in these more general coordinates is revealed to be a significant change, requiring the metric tensor, because the former is a covariant vector and the other contravariant. To transform such a steepest descent update to the r system, we individually transform the starting vector and the gradient, whose components are of contravariant and covariant types respectively:

$$r_0^\mu = (t^{-1})^\mu_\nu s_0^\nu, \;\; \Phi_{,\mu}(r_0) = t^\nu_\mu \Phi_{,\nu}(s_0). \tag{11}$$

These two quantities are then combined to produce an update in the r system:

$$r_1^\mu = r_0^\mu + \Delta r^\mu = r_0^\mu - \beta\, g^{\mu\lambda}\Phi_{,\lambda}(r_0), \tag{12}$$

and the resulting contravariant vector r_1^μ can then be transformed back to the s system:

$$s_1^\mu = t^\mu_\nu r_1^\nu. \tag{13}$$

The s_1^μ vectors in (10) and (13) are not in general the same.* Thus steepest descent optimizations in any pair of coordinate systems related through t^μ_ν should be expected to produce different convergence histories.

Transformation of Hessian matrices

A quadratic objective function has ellipsoidal isosurfaces.† Unequal diagonal elements of $\phi_{\mu\nu}$ affect ellipticity, and non-zero off-diagonal elements affect ellipticity and produce axial re-orientations. The Hessian matrix changes under t^μ_ν according to

$$\phi_{\mu\nu}(r) = t^\lambda_\mu\, \phi_{\lambda\sigma}(s)\, t^\sigma_\nu. \tag{14}$$

*It may appear that we are just performing the same tasks in different domains, but the way displacements and direction vectors are mixed in (12) causes a significant change. In fact, if the values of Φ arrived at in the separate line searches in (10) and (12) are different (and nothing but coincidence could make them the same), by (4) it is impossible for the two post-update positions in the s system to agree.

†An isosurface of Φ is the locus of s^μ vectors whose Φ values are equal. A contour on a topographic map is a 2D isosurface.

Inspection of this relationship reveals that differences between the diagonal elements of t^{μ}_{ν} cause changes in the ratios of the diagonal elements of the Hessian after transformation, affecting ellipticity; non-zero off-diagonal elements in t^{μ}_{ν} introduce relative changes in both the on- and off-diagonal elements of the Hessian after transformation, affecting ellipticity and axis orientation. Consequently, suitably general coordinate transforms t^{μ}_{ν} can alter almost all shape parameters of a quadratic objective function.

DESIGNING TRANSFORMS BASED ON THE HESSIAN MATRIX

In (4) we observe that by finding the minimum of a quadratic objective function in one coordinate system we find it in all coordinate systems. Comparison of (10) and (13) confirms that although gradients map back and forth uniquely between coordinate systems, steepest descent updates do not, and for a given starting point each such update is special to its coordinate system. From (7) and (8) we observe that the gradient and the Newton update are parallel in a system in which $\phi_{\mu\nu} = \delta_{\mu\nu}$. Finally, inspection of (14) suggests that transformations can be designed to produce almost any desired change in the ellipsoidal parameters of isosurfaces in $\phi_{\mu\nu}$.

Formulation of the problem

From these facts it follows that in a system characterized by the Hessian $\phi_{\mu\nu}$, we should be able to design transforms t^{μ}_{ν} such that $t^{\lambda}_{\mu}\,\phi_{\lambda\sigma}\,t^{\sigma}_{\nu} = \delta_{\mu\nu}$, carry out an efficient steepest-descent based search for the minimum in this system, and then transform the result back to our original (and presumably otherwise more appropriate) coordinate system. Because $\phi_{\mu\nu}$ is symmetric, in an N dimensional problem the equations

$$t^{\lambda}_{\mu}\,\phi_{\lambda\sigma}\,t^{\sigma}_{\nu} = \delta_{\mu\nu} \tag{15}$$

are $N(N+1)/2$ in number, which means $N(N-1)/2$ further degrees of freedom remain to be fixed when choosing t^{μ}_{ν}. There are therefore many transforms to choose from which satisfy (15), and other considerations can be invoked in making a specific selection.

Approach

There are many possible ways to reduce the remaining degrees of freedom and precisely specifying a particular t^{μ}_{ν}. For instance, we could fill $N(N-1)/2$ of the columns of t^{μ}_{ν} with basis vectors which span a portion of model space in some attractive way, after which (15) would constrain the others. Or, we could pre-select the lower triangular entries of t^{μ}_{ν}, and treat them as fixed while the other entries are determined. This second approach is in some sense the most general, in that no one basis vector of the new coordinate system is entirely pre-selected, and no one basis vector is entirely determined by (15). The left-most columns are almost entirely pre-selected, and the right-most almost entirely determined, and one might object that this reduces the generality of the approach. But since the model vector can usually be re-ordered without changing the problem, this is an apparent effect only.

NUMERICAL CONSIDERATIONS

In a companion report (Innanen, 2020d) numerical procedures for computing general $N \times N$ transformation matrices satisfying (15), given the $N(N-1)/2$ lower-triangular entries as input, are set out. The algorithm moves from left to right through the columns of t^{μ}_{ν}; at the jth column the j unknowns at and above the diagonal are determined.

Involvement of the Hessian

In large optimization problems, the Hessian matrix is generally too large and complex to be calculated, stored, or inverted. At most, Truncated Newton type methods are applied, which involve Hessian information, but only in the form of Hessian-vector products, which are computationally tractable. In designing the transformation matrix, all calculations similarly involve Hessian-vector products only, and so no Hessian information beyond what is generally involved in Truncated Newton methods appears to be required.

Accumulating sets of matrix row operations

At the jth column of t^{μ}_{ν}, the unknown on the diagonal is the solution of a quadratic equation in one unknown, and the elements above the diagonal are the solutions of linear equations, each in one unknown, calculations which are trivial computationally. The weights within these equations are, however, determined through sequences of row operations on a matrix built from Hessian-vector products. The jth column itself produces on the order of j of these row operations, and several elements of the matrix must be subject to all previous row operations at each j. These two sets of operations represent the bulk of the computational expense involved in the design of t^{μ}_{ν}. Current codes for determining t^{μ}_{ν} do not handle these repeated calculations particularly efficiently. However, matrix row operations being a very common numerical operation, it is likely methods exist which make this process quite efficient.

CONCLUSIONS

Re-parameterizations of model space, which are widely applied in seismic inverse problems, are coordinate transforms between non-Cartesian systems. Under a class of transformations constrained by the Hessian operator in the reference system, optimization should have favourable convergence properties. This class of transforms includes an infinite number of variants, and seeking examples from within this class with other, additional, favourable features appears warranted. Currently the degree to which the computational expense of determining the transformation operator can be reduced is not fully clear; this will be an important question to resolve.

REFERENCES

Innanen, K. A., 2020a, Rep 1. A review of tensors in non-cartesian coordinate systems: CREWES Annual Report, **32**.

Innanen, K. A., 2020b, Rep 3. Application of misfit-based model space coordinate system design to seismic AVO inversion: CREWES Annual Report, **32**.

Innanen, K. A., 2020c, Rep 4. Application of misfit-based model space coordinate system design to seismic full waveform inversion: CREWES Annual Report, **32**.

Innanen, K. A., 2020d, Rep 5. Numerical procedures for computing constrained coordinate transforms: CREWES Annual Report, **32**.