Convex Duality and Separable Inverse Problems in Wave Propagation

William W. Symes*

Abstract

A number of inverse problems in wave propagation admit division of the model space into *kinematic* or propagator variables and *dynamic* or amplitude (or reflectivity, or source, or...) variables. The dynamic variables have a linear or essentially linear influence on the simulated data, hence the best fit formulations of inverse problems are *separable* in the sense of nonlinear programming. Mean square objective functions for the analysis of such problems may be studied either in their original, primal form or, *via* the Fenschel-Rockefellar Theorem of convex analysis, in their dual form. The duality principle displays connections between apparently unrelated approaches: for example the semblance function of Migration Velocity Analysis in reflection seismology is dual to the Output Least Squares objective, under certain circumstances. An abstract setting allows a uniform approach to the study of a wide variety of objective functions for waveform inversion. For a subset of these objectives (the differential semblance class) the duality relation permits a simple proof of smoothness of the objective as function of the kinematic variables, and suggests an accurate algorithm for computation of approximate gradients. For a still smaller subset of objective functions and problems, it is possible to demonstrate that only one critical point exists in a large domain of kinematic variables. For such problems and formulations, global solution of the inverse problem is possible using local smooth optimization techniques.

Author's Note: These notes form a draft of joint work in progress with Prof. Guy Chavent. I wish to record my indebtedness to Prof. Chavent for provision of the key insight, for his collaboration in working out its implications, and for his hospitality, and that of Université de Paris IX and INRIA Roquencourt, where this work was carried out.

*The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston TX 77251-1892

Introduction

Quite a few forward modeling operators for simulation of wave propagation have the form

$$(s, r) \mapsto B[s]r$$

Here the *model space* is represented as a Cartesian product $(s, r) \in S \times R$, and the forward modeling or simulation operator B as a family of linear operators on the Hilbert space R of *reflectivities* parameterized by an open subset S of *propagators*. The diction here is borrowed from the acoustic or elastic model of primaries only reflection seismograms, but actually applies to a much wider class of problems. All have in common that the propagators actually amount to a set of parameters for wave velocities, i.e. determining kinematics, and the reflectivities are a complementary set of variables determing dynamics - these latter can be reflectivities in the sense of reflection modeling, i.e. wavelength scale fluctuations in elastic parameters, or even source parameters, as I shall illustrate below. In any case the dynamic variables have a linear influence on the simulated data, whereas the propagators typically do not.

Several years ago I introduced an approach to estimation of the entire model (s, r) for problems of this type, through a variational principle combining event semblance and data fit measures. For several simple problems, proof had emerged for excellent global properties of this *differential semblance* principle, suggesting that efficient gradient based numerical local optimization algorithms could be used to estimate both propagators and reflectivities directly from waveform data, without the benefit of an accurate initial propagator. Numerical experiments have supported this conclusion. So far as I know, differential semblance ("DSO") is the only approach to velocity estimation to exhibit this ability.

DSO is but one of many optimization formulations of inverse problems in seismology advanced over the last decade, and the question naturally arose as to its relation to other approaches. The mathematical reasons for its relative success have been obscure until recently, despite the theoretical and computational results already mentioned. The calculation of the gradient of the differential semblance objective function appear to demand a rather complex approach ([5]) and moreover seems quite expensive, requiring an *ab initio* iterative solution of a large linear system for each evaluation. To complete this catalogue of aches and pains, the objective function depends on a penalty parameter, and no rational method appeared to be available to determine this quantity.

Some of this fog has been lifted by the work of Hua Song and Mark Gockenbach, presented elsewhere in the TRIP 1994 Annual Report. Prof. Guy Chavent ([1]) has provided an enormously important insight, that the *dual* to the DSO objective function, in the sense of convex analysis, defines an interesting optimization problem in its own right with possibly superior properties. In the following page I have sketched a synthesis of these ideas. The result is an alternative to the original formulation of DSO with the following features:

- The gradient calculation computation is much simplified: the new dual approach achieves control over the gradient accuracy without the theoretical and computational complexity of the primal approach ([5]);

- The dual formulation allows the nonlinear iterations for the propagator variable s to be interleaved with linear iterations for the reflectivity variable r, which should result in tremendous computational savings (this is pointed out already in [1]);

- The dual formulation suggests a *scaling* of the differential semblance operator, related to Chavent's idea of *quantitative migration*, which yields a much simpler problem for the the inner variable r. In fact it should be possible to produce estimates for r, using Kirchhoff methodology for simulation and related processes, without the use of an iterative method, or with at most a small number of iterations of a very well-preconditioned Lanczos process;

- The scaling mentioned above has as a happy consequence, noted already by [6], that the resulting modified DSO objective function appears to a much larger domain of convexity, or at any rate one can prove convergence to a globally satisfactory solution under much less restrictive hypotheses on the initial estimate.

The upshot appears to be a complete - and very favorable - revision of the economics of DSO.

I will set out the abstract machinery of dual DSO here in note form. I work out the implications in detail only for the very simplest version of the transmission inverse problem, and without any numerical tests. It will be evident that the same ideas apply to more complex setting, particularly reflection inversion.

It is interesting to note that duality is usually of mostly theoretical interest in optimization. In this application, however, it appears to promise returns considerably beyond improvement in theory.

Abstract Setting for Multiexperiment Inverse Problems

Optimization Formulations of Separable Inverse Problems

Set of simulation parameters ("propagators"): S

In order that calculus operations be well-defined, S must be an open set in a Banach manifold.

Reflectivity model Space R (separable Hilbert space)

Semblance operator $W : R \to F$, F another Hilbert space.

A useful family of norms on R: for $\epsilon > 0$,

$$\|r\|_\epsilon^2 \equiv \|W\|_F^2 + \epsilon^2\|r\|_R^2$$

Data space D - Hilbert space.

Simulation operator $B : S \to L(R, D)$

Given $d \in D$:

Primal formulation: minimize over $s \in S, r \in R$

$$J_\sigma[s,r] \equiv \frac{1}{2}(\|B[s]r - d\|_D^2 + \sigma^2\|r\|_\epsilon^2)$$

This formulation accomodates regularization and a penalty for the semblance operator W in the last term.

Dual formulation: minimize over $s \in S, q \in D$

$$J_\sigma^*[s,q] \equiv \frac{1}{2}\left(\frac{1}{\sigma^2}\|B[s]_\epsilon^* q\|_\epsilon^2 - 2 < q, d >_D + \|q\|_D^2\right)$$

Reduced objectives:

$$J_\sigma^{\rm red}[s] = \inf_{r \in R} J_\sigma[s,r]$$

$$J_\sigma^{\rm red,*}[s] = \inf_{q \in D} J_\sigma^*[s,q]$$

Fenschel-Rockefellar Duality Theorem:

$$J_\sigma^{\rm red}[s] + J_\sigma^{\rm red,*}[s] \equiv 0$$

Gockenbach ([4]) observed that the behaviour of $J_\sigma^{\rm red}$ for small σ is the critical feature of DSO, which explains most of its important properties. A good scaling for studying the limit $\sigma \to 0$: divide $J_\sigma^{\rm red,*}$ and q by σ^2 to get

$$\frac{1}{\sigma^2} J_\sigma^{\rm red,*}[s] \equiv \inf_{q \in D} \frac{1}{2}(\|B[s]_\epsilon^* q\|_\epsilon^2 - 2 < q, d >_D + \sigma^2\|q\|_D^2)$$

Normal equations (with $\sigma \to 0$ scaling): for primal formulation,

$$(B_\epsilon^*[s]B[s] + \sigma^2 I)r = B_\epsilon^*[s]d$$

solution denoted by $r_\sigma[s]$, and for the dual formulation

$$(B[s]B_\epsilon^*[s] + \sigma^2 I)q = d$$

solution denoted by $q_\sigma[s]$.

Duality relations:

$$r_\sigma[s] = B_\epsilon^*[s]q_\sigma[s]$$
$$\sigma^2 q_\sigma[s] = d - B[s]r_\sigma[s]$$

The $\sigma \to 0$ limits are well-defined only when $d \in \text{Range}B[s]$:

$$J_0^*[s] \equiv \lim_{\sigma\to 0}\frac{1}{\sigma^2}J_\sigma^{\text{red},*}[s]$$
$$= -\frac{1}{2} < q_0[s], d >_D$$

where $q_0[s]$ is any solution of

$$B[s]B_\epsilon^*[s]q = d$$

Similarly

$$J_0[s] \equiv \lim_{\sigma\to 0}\frac{1}{\sigma^2}J_\sigma^{\text{red}}[s]$$
$$= \inf\{\frac{1}{2}\|r\|_\epsilon : B[s]r = d\}$$

A Setting for Multiexperiments

Basic model spaces: Gel'fand Triple $V \subset H \subset V'$.

Set of experiment parameters (measure space): $E, d\mu$

Maximal model space: $R^0 = L^2(E, H, d\mu)$

Model space (to be chosen): $R \subset L^2(E, V, d\mu)$

Coherent model subspace: $R_0^0 \equiv \{u \in R^0 : u(e_1) = u(e_2) \text{ for almost all } e_1, e_2 \in E\}$

Semblance operator: $W : R \subset L^2(E, V, d\mu) \to R^0$ for which $\ker W = R_0^0 \bigcap R \equiv R_0$. The domain R should be dense in $L^2(E, V, d\mu)$.

Families of norms on V and H, parameterized by E: $\|(\cdot)\|_{V,e}$, $\|(\cdot)\|_{H,e}$. May all be same, in any case are all equivalent:

$$C_*\|u\|_{V,e_0} \le \|u\|_{V,e_1} \le C^*\|u\|_{V,e_0}$$

for any $e_0, e_1 \in E$, suitable $0 < C_* < C^*$, and similarly for H.

Family of norms on R (essentially the graph norm of W):

$$\|u\|_\epsilon^2 = \int_E d\mu(e)[\|Wu(e)\|_{H,e}^2 + \epsilon^2\|u(e)\|_{V,e}^2]$$

These define an operator $N_\epsilon : R \to R'$ *via*

$$< u, N_\epsilon v > \equiv < u, v >_\epsilon$$

By Lax-Milgram N_ϵ is invertible.

Basic data space: T.

Multiexperiment data space: $D \equiv L^2(E, T, d\mu)$

Simulator:

$$\begin{aligned} B : S &\to L^\infty(E, L(V, T), d\mu) \\ &\subset L(R, D) \end{aligned}$$

the latter inclusion being the natural one.

Pointwise adjoint:

$$\begin{aligned} B' : S &\to L^\infty(E, L(T, V'), d\mu) \\ &\subset L(D, R') \end{aligned}$$

Hilbert space adjoint of $B[s]$ as map: $(R, \|(\cdot)\|_\epsilon) \to (D, \|(\cdot)\|_D)$:

$$B^*_\epsilon[s] = N_\epsilon^{-1} B'[s]$$

Possibility that the norms $\|(\cdot)\|_{V,\epsilon}$ depend on $s \in S$ is alllowed as well. Will take advantage of this later.

Examples based on Transmission Inversion

This problem has been studied in [6], who showed the smoothness of the DSO objective function and its good global properties (see [6] and below for details).

Experiment space: $E = [Z_{\min}, Z_{\max}]$ = interval of source depths in source well. Measure = Lesbesgue.

For simplicity assume single receiver in receiver well.

Simulation parameters: a set S of slowness distributions $s(x, z)$ in the region between the wells. Restrictions: 1) smoothness, 2) only generate ray fields without caustics between wells.

Source waveform: $r(t)$. Key assumption: source waveform is independent of source position $z_s \in E$. Also assume isotropy for convenience.

Forward map, approximated by geometric optics under the assumption of unique source - receiver rays:

$$B[s]r(z_s,t) = \phi(t)a[s](z_s)r(t-\tau[s](z_s)),\ z_s \in E,\ t \in [0,T]$$

where: $\tau[s](z_s)$ = geometric optics traveltime or geodesic distance function from source at depth z_s to receiver, $a[s](z_s)$ = corresponding amplitude, and $\phi(t) \in C_0^\infty(0,T)$ is a cutoff or window function.

Still to be chosen: spaces for s, r, data. Will discuss two possibilities in detail.

In both approaches it's convenient to choose S to be a subset of $C^\infty(\mathbf{R}^2)$ parameterized by the vector $c \in C \subset \mathbf{R}^{N_s}$, as follows:

$$s(x,z) = s_0 \exp\left(\sum_{i=1}^{N_s} c_i \psi_i(x,z)\right)$$

where $\{\psi_i : i = 1, ..., N_s\} \subset C_0^\infty$ is a set of smooth basis or shape functions and s_0 is an ambient slowness level. The coordinate neighborhood C must be chosen so that the slownesses in S generate ray families without caustics, at least for rays connecting the sources and receivers.

In this way the meaning of "differentiable" for functions of s is obvious. With somewhat more technical difficulty one could accomodate infinite dimensional families of sufficiently smooth functions, but this seems pointless.

Coherent Formulation: Output Least Squares

Selection of R: Take

$$V = H = V' = L^2([T_{\min}, T_{\max}])$$

Family of norms parameterized by E: determined by a positive function $w \in L^\infty(E)$, as follows: for $u \in V$, $z_s \in E$,

$$\|u\|^2_{V,z_s} \equiv w(z_s) \int_{T_{\min}}^{T_{\max}} dt |u(t)|^2$$

$$R = R^0 = L^2(E,H) \simeq L^2([Z_{\min}, Z_{\max}] \times [T_{\min}, T_{\max}])$$

Regard functions in H as *periodic in* t, and interpret $t - \tau[s](z_s)$ in definition of $B[s]$ as *periodic shift.*

Coherency operator:

$$Wr(e) = r(e) - \frac{1}{\mu(E)} \int_E d\mu(e')r(e')$$

Data Space: $T = L^2[0,T], D = L^2(E,T) \simeq L^2([Z_{\min}, Z_{\max}] \times [0,T])$

Primal formulation: minimize over $s \in S, r \in R$

$$J_\sigma[s,r] = \frac{1}{2}[\|B[s]r - d\|_D^2 + \sigma^2 \|r\|_\epsilon^2]$$

Output Least Squares, usual formulation: limit of $\sigma \to \infty$, $\epsilon \to 0$ so that $\sigma\epsilon \equiv \delta$ is constant. Limit problem is: minimize over $s \in S, r \in R_0 \simeq H$

$$J_{\rm OLS}[s,r] = \frac{1}{2}(\|B[s]r - d\|^2 + \delta^2 \|r\|_H^2)$$

Dual Formulation with original (unscaled) variable: minimize over $s \in S,\ q \in D$

$$J_\sigma^*[s,r] = \frac{1}{2}(\frac{1}{\sigma^2}\|B_\epsilon^*[s]\|^2 - 2 < q, d > + \|q\|^2)$$

Once again the same limit as described above gives the dual formulation to Output Least Squares: minimize

$$J_{\rm OLS}^*[s,r] = \frac{1}{2}(\frac{1}{\delta^2}\|B_0^*[s]q\|^2 - 2 < q, d > + \|q\|^2)$$

where

$$B_0^*[s]q = \frac{1}{|E|}\int_E dz_s B'[s](z_s)q(z_s)$$

A very important specialization occurs when the norm in R has been chosen so that the pointwise dual B' is a *quantitative migration operator*, i.e.

$$B'[s](e)B[s](e) \approx I$$

for each $e \in E$. This concept was introduced in [2]; it makes somewhat more precise the notion of *true amplitude migration* used in the reflection seismology literature.

For the present problem, this property results from the use of the experiment- (z_s-) and propagator-dependent norm in $H \simeq H \times \{z_s\}$ given by choosing $w(z_s) = a[s]^2(z_s)$ in the above formulas. Then for $u \in V$, $p \in T$

$$< B[s](z_s)u, p >_T = \int_0^T dt \phi(t) a[s](z_s) u(t - \tau[s](z_s)) p(t)$$

$$= \int_{T_{\rm min}}^{T_{\rm max}} dt a[s](z_s) u(t) (\phi p)(t + \tau[s](z_s))) = < u, B[s]'(z_s)p >_{H,z_s}$$

whence for $q \in D$

$$B'[s](z_s)q(z_s,t) = a[s]^{-1}(z_s)(\phi q)(z_s, t + \tau[s](z_s))$$

Thus

$$B_0^*[s]B[s]r(t) = \frac{1}{|E|}\left(\int dz_s \phi^2(t+\tau[s](z_s))\right) r(t)$$
$$= r(t)$$

so long as supp $r \subset \{t : \phi(t+\tau(z_s)) \equiv 1\}$. Note that for any such r,

$$\|B[s]r\| = \|r\|$$

From the Duality Relations above applied to the OLS problem and its dual, the quantities r_δ^0 and q_δ^0 which extremize the two OLS objective and its dual satisfy

$$\delta^2 r_\delta^0 = B_0^*[s]q_\delta^0,\;\; q_\delta^0 = d - B[s]r_\delta^0$$

whence

$$r_\delta^0 = (B_0^*[s]B[s] + \delta^2 I)^{-1}B_0^*[s]d \approx \frac{1}{1+\delta^2}B_0^*[s]d$$

if d satisfies supp $d(z_s,\cdot) \subset \{t : \phi(t) = 1\}$ which we assume from now on. Then

$$q_\delta^0 \approx (1-(1+\delta^2)^{-1}B[s]B_0^*[s])d$$

so

$$\|B_0^*[s]q_\delta^0\|^2 = \|B_0^*[s]d - (1+\delta^2)^{-1}B_0^*[s]B[s]B_0^*[s]d\|^2$$
$$\approx \left(\frac{\delta^2}{1+\delta^2}\right)^2 \|B_0^*[s]d\|^2$$

etc. Eventually find

$$\inf_{q\in D} J_{OLS}^*[s,q] \approx \frac{1}{2}\left(\frac{1}{1+\delta^2}\|B_0^*[s]d\|^2 - \|d\|^2\right)$$

Thus according the Fenschel-Rockefellar theorem instead of the minimizing the output least squares one could just as well maximize

$$\|B_0^*[s]d\|^2 = \text{const.} \int dt \left(\int_E dz_s d(z_s, t+\tau[s](z_s))\right)^2$$

This functional is the equivalent for the transmission inversion problem of the *semblance function* of Migration Velocity Analysis in reflection seismology.

Incoherent Formulation: Differential Semblance

Selection of R:

$$V = \{u \in H^1[T_{\min}, T_{\max}] : u(T_{\min}) = u(T_{\max}), \int_{T_{\min}}^{T_{\max}} u = 0\}$$

(viewed as *periodic* $H^1[T_{\min}, T_{\max}]$ functions with zero mean). Norm on V parameterized by $w \in L^\infty[Z_{\min}, Z_{\max}]$:

$$\|u\|_{V,z_s} \equiv w(z_s) \int_{T_{\min}}^{T_{\max}} \left|\frac{\partial u}{\partial t}\right|^2$$

Similarly,

$$H = L^2[T_{\min}, T_{\max}]$$

with

$$\|u\|_{H,z_s} \equiv \int_{T_{\min}}^{T_{\max}} |u|^2$$

(i.e. in H the norm is *unweighted*). Note that

$$\sqrt{w}\frac{d}{dt} : V \to H$$

is an isometric isomorphism.

Coherency operator:

$$W = \sqrt{w}\,\frac{\partial}{\partial z_s} : R \to R^0$$

The model space $R \subset L^2(E, V)$ (domain of W) is

$$R = H^1[E, H] \bigcap L^2[E, V]$$

$$\simeq \{u \in H^1([Z_{\min}, Z_{\max}] \times [T_{\min}, T_{\max}]) : u(z_s, T_{\min}) = u(z_s, T_{\max}) \text{ and}$$

$$\int_{T_{\min}}^{T_{\max}} dt u(z_s, t) = 0 \text{ for almost all} z_s \in [Z_{\min}, Z_{\max}]\}$$

Family of norms in R defined as usual by

$$\|r\|_\epsilon^2 \equiv \int_E dz_s (\|Wr\|_{H,z_s}^2 + \epsilon^2 \|r\|_{V,z_s}^2)$$

$$= \int_E dz_s \left(\left\|\sqrt{w}\frac{\partial r}{\partial z_s}\right\|_H^2 + \epsilon^2 \left\|\sqrt{w}\frac{\partial r}{\partial t}\right\|_H^2 \right)$$

Operator $N_\epsilon : R \to R'$ defined by these norms is

$$N_\epsilon = \left(\sqrt{w}\frac{\partial}{\partial z_s}\right)' \left(\sqrt{w}\frac{\partial}{\partial z_s}\right) + \epsilon^2 \left(\sqrt{w}\frac{\partial}{\partial t}\right)' \left(\sqrt{w}\frac{\partial}{\partial t}\right)$$

formally equivalent to the elliptic operator

$$-\left(\frac{\partial}{\partial z_s} w \frac{\partial}{\partial z_s} + \epsilon^2 \frac{\partial}{\partial t} w \frac{\partial}{\partial t}\right)$$

on the model rectangle with periodic boundary condition in t and Neumann conditions in z_s.

To better "balance" the simulator with the coherency operator, redefine forward map:

$$B[s]r(z_s,t) = \phi(t)a[s](z_s)\frac{\partial r}{\partial t}(z_s, t-\tau[s](z_s))$$

Thus $r \in R$ now represents the (*a priori* location dependent) *primitive* of the actual source waveform.

The $\sigma \to \infty$ limit gives a variant of the Output Least Squares objective, just as in the coherent case. The $\sigma \to 0$ limit is more interesting, and will be studied extensively below: in terms of the primal variables,

$$J_0[s] = \inf\{\frac{1}{2}\left(\left\|\sqrt{w}\frac{\partial r}{\partial z_s}\right\|^2_{L^2(E,H)} + \epsilon^2\left\|\sqrt{w}\frac{\partial r}{\partial t}\right\|^2_{L^2(E,H)}\right) : B[s]r = d\}$$

Evidently this is the analog of the dual Output Least Squares objective. When B' is a quantitative migration, so that one can replace r by $B^*_\epsilon[s]d$ in the above formula, it becomes a *differential* analogue of the semblance function from Migration Velocity Analysis introduced in the transmission context above, whence the name of this approach.

Smoothness and Gradient Calculation

Generalities

Duality provides a particularly simple avenue to proof of smoothness of the reduced objective functions. This is so because the normal equations for the dual reduced amplitude variable $q_\sigma[s]$ reads

$$(B[s]B^*_\epsilon[s] + \sigma^2 I)q_\sigma[s] = d$$

so that $q_\sigma[s]$ is smooth in s if the same is true of the *dual normal operator* $M_\epsilon[s] \equiv B[s]B^*_\epsilon[s] = B[s]N_\epsilon^{-1}B'[s]$. Since

$$J_\sigma^{\mathrm{red},*}[s] = -\frac{1}{2} < q_\sigma[s], d >$$

smoothness of both reduced objectives follows immediately from smoothness of the dual normal operator and the Fenschel-Rockefellar theorem. The same cannot be said of the primal reduced amplitude variable $r_\sigma[s]$, for which the normal equations read

$$(B^*_\epsilon[s]B[s] + \sigma^2 I)r_\sigma[s] = B[s]^*_\epsilon d$$

Here smooth dependence on s of the operator on the left hand side is insufficient to render r smooth. The right hand side must also be smooth, and as we shall see in examples this is an extremely restrictive assumption.

In many cases the forward map may be factored as

$$B[s] = \Phi[s]\tilde{B}[s]$$

where $\tilde{B}[s] : R \to \tilde{D}$ is invertible and $\Phi[s] : \tilde{D} \to D$ is smooth in s; here $\tilde{D}$ is a suitable intermediate space. Then the dual normal operator is smooth if

$$\tilde{B}[s] N_\epsilon^{-1} \tilde{B}[s]'$$

is smooth, which in turn is true if

$$\tilde{B}[s]'^{-1} N_\epsilon \tilde{B}[s]^{-1}$$

is smooth. This operator is ofter easier to prove smooth.

Computation of the Gradient

We suppose that some sort of Riemann metric has been established on the manifold S of propagators, say $< \cdot, \cdot >_S$. The the gradient of the reduced objective is defined by

$$< \mathrm{grad} J_\sigma^{\mathrm{red}}[s], \delta s >_S = D J_\sigma^{\mathrm{red}}[s] \delta s$$

Since an alternate expression for the reduced objective function (thanks to duality) is

$$J_\sigma^{\mathrm{red}}[s] = \frac{1}{2} < q_\sigma[s], d >$$

$$= \frac{1}{2} < (M_\epsilon[s] + \sigma^2 I)^{-1} d, d >$$

it follows that

$$D J_\sigma^{\mathrm{red}}[s] \delta s = -\frac{1}{2} < (M_\epsilon[s] + \sigma^2 I)^{-1} (D M_\epsilon[s] \delta s)(M_\epsilon[s] + \sigma^2 I)^{-1} d, d >$$

$$= -\frac{1}{2} < (D M_\epsilon[s] \delta s) q_\sigma[s], q_\sigma[s] >$$

$$= -\frac{1}{2} \{ 2 < (DB[s]\delta s) r_\sigma[s], q_\sigma > - < r_\sigma, (D N_\epsilon[s] \delta s) r_\sigma > \}$$

Define operators $s \in S \mapsto H_B[s] \in M(R, D, T_s S)$ ($M(H_1, H_2, ..., H_n, K)$ = continuous n-multilinear maps on $H_1 \times H_2 \times ... \times H_n$ with values in K) by

$$< (DB[s]\delta s) r, q > = < \delta s, H_B[s](r, q) >_S$$

and similarly $s \in S \mapsto H_N[s] \in M(R, R, T_s S)$ by

$$< (DN[s]\delta s) r_1, r_2 > = < \delta s, H_N[s](r_1, r_2) >_S$$

Then

$$\mathrm{grad}\, J_\sigma^{\mathrm{red}}[s] = -H_B[s](r_\sigma[s], q_\sigma[s]) + \frac{1}{2} H_N[s](r_\sigma, r_\sigma)$$

A similar expression may be derived from the primal formulation.

Several remarks are in order.

Remark 1 The operators H_B may be computed *via* the adjoint state technique when B is computed by solving a boundary value problem - this is done for example in [5] for the acoustic wave equation. For other simulators, eg. Kirchhoff, it's even simpler.

Remark 2 Depending on definition of the norms in V and H, H_N might also be computed *via* the adjoint state technique or by an explicit integral formula. In some cases N is independent of S and this term drops out (cf. [5] for example).

Remark 3 Generally $q_\sigma[s]$ cannot be computed exactly, but must be approximated: i.e. we have available an *estimate* $\tilde{q}_\sigma[s]$. This estimate might be generated either by applying an iterative method to the dual normal equations, or by applying an iterative method to the primal normal equations to estimate $\tilde{r}_\sigma[s]$ followed by use of the second duality relation to obtain $\tilde{q}_\sigma[s] = \sigma^{-2}(d - B[s]\tilde{r}_\sigma[s])$. *In this case it is extremely important to replace the formula above by*

$$\mathrm{grad}\, J_\sigma^{\mathrm{red}}[s] = -H_B[s](N_\epsilon^{-1}[s]B[s]'\tilde{q}_\sigma[s], \tilde{q}_\sigma[s]) + \frac{1}{2}H_N(N_\epsilon^{-1}[s]B[s]'\tilde{q}_\sigma[s], N_\epsilon^{-1}[s]B[s]'\tilde{q}_\sigma[s])$$

This is because the first term in the formula for the directional derivative of J_σ^{red} is actually

$$\frac{1}{2}(< \tilde{q}_\sigma[s], [(DB[s]\delta s)N_\epsilon^{-1}[s]B[s]'\tilde{q}_\sigma[s] + B[s]N_\epsilon^{-1}[s](DB[s]\delta s)']\tilde{q}_\sigma[s] >$$

The operator appearing in this expression is symmetric, i.e. the symmetric part of

$$(DB[s]\delta s)N_\epsilon^{-1}B[s]'$$

This operator is typically unbounded in the examples, whereas its symmetric part is bounded. In effect the use of an unsymmetric approximation to the this symmetric expression, for example

$$H_B[s](\tilde{r}_\sigma[s], \tilde{q}_\sigma[s])$$

tends to create a very large error in the gradient.

Remark 4 Similar difficulties arise in devising computable approximations to the gradient expression derived from the primal formulation. This topic was considered in detail in [5], which explained and demonstrated an accurate approach based on polynomial iteration for $\tilde{r}_\sigma[s]$. As observed above, the form of the primal normal equations indicates that $r_\sigma[s]$ does not have a bounded derivative, so an accurate gradient expression necessarily involves an approximation of this derivative, to avoid arbitrarily large errors. This reasoning leads to the addition of a "correction" term, which vanishes when $\tilde{r} = r$ and causes the error in the gradient to be dominated by size of the normal residual. Unfortunately this latter conclusion appears to require the use of a *static* polynomial iteration, for example the Chebyshev method, as discussed in [5], rather than more efficient *dynamic* iterations such as the conjugate gradient or other Lanczos-type methods. The approximation based on the dual formulation, as explained above, carries no such restriction.

Example: Transmission Inversion

Coherent Formulation: Necessity of Data Smoothness

Using the z_s-dependent norm that makes the adjoint B_0^* into a quantitative migration operator, as discussed above, one sees that the r component of the output least square solution is

$$r_{OLS}[s](t) = \frac{1}{|E|}\int_E dz_s d(z_s, t+\tau[s](z_s))$$

$$= \frac{1}{|E|}\int_E dz_s r^*(t+\tau[s](z_s) - \tau^*(z_s))$$

for exact data $d = B[s^*]r^*$. Note that the RHS of the primal normal equations is not smooth unless d is smooth, as claimed above. Also

$$q_{OLS}[s](z_s,t) = d(z_s,t) - \frac{1}{|E|}\int_E dz_s' r^*(t+\tau[s](z_s') - \tau[s](z_s) - \tau^*(z_s'))$$

Smoothness of $q_{OLS}[s]$, hence of $J^{\rm red}_{OLS}[s]$, clearly requires that r^* be smooth. The same is easily established for the full reduced objective $J^{\rm red}_{\sigma[s]}$.

Also easy to see: $M_\epsilon[s]$ is not a smooth operator valued function of s.

Conditions for Smoothness of Differential Semblance Objective

As suggested above write

$$B[s] = \Phi\tilde{B}[s]$$

where

$$\Phi : S \to L(R,D),\ \Phi[s]r = \phi a[s]\frac{\partial r}{\partial t}$$

$$\tilde{B}[s] \equiv T_{-\tau[s]} : R \to R,\ T_{-\tau[s]}r(z_s,t) = r(z_s, t-\tau[s](z_s))$$

Observe that Φ is smooth in s because this is true of linear differential operators depending on parameters. Also note that $T_{-\tau[s]}$ is actually a bounded and boundedly invertible linear operator on R, as follows for example from the finite difference definition of H^1 ([3]), and $T^{-1}_{-\tau[s]} = T_{\tau[s]}$. Same is true for the pointwise adjoints $T'_{-\tau[s]} : R' \to R'$.

Commutation of $T_{\pm\tau}$ with differentiation: define $T^0_{\pm\tau} : R^0 \to R^0$ by the same formula as defines $T_{\pm\tau}$ but now regarded as an operator on $R^0 = L^2(E,H)$. Note that $T^0_{\pm\tau}$ is a family of unitary maps, with $(T^0_{\pm\tau})' = (T^0_{\pm\tau})^{-1} = T^0_{\mp\tau}$. It is easy to see that

$$\frac{\partial}{\partial z_s}T_{-\tau} = T^0_{-\tau}\left(\frac{\partial}{\partial z_s} - \frac{\partial \tau}{\partial z_s}\frac{\partial}{\partial t}\right)$$

$$\frac{\partial}{\partial t}T_{-\tau} = T^0_{-\tau}\frac{\partial}{\partial t}$$

Now

$$(\tilde{B}[s]')^{-1}N_\epsilon(\tilde{B}[s])^{-1} = T_\tau'\left(\left(\frac{\partial}{\partial z_s}\right)' w[s]\left(\frac{\partial}{\partial z_s}\right) + \epsilon^2\left(\frac{\partial}{\partial t}\right)' w[s]\left(\frac{\partial}{\partial t}\right)\right)T_\tau$$

$$= \left(\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right)'(T_\tau^0)'w[s]T_\tau^0\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right) + \epsilon^2\left(\frac{\partial}{\partial t}\right)'(T_\tau^0)'w[s]T_\tau^0\left(\frac{\partial}{\partial t}\right)\right)$$

$$= \left(\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right)' w[s]\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right) + \epsilon^2\left(\frac{\partial}{\partial t}\right)' w[s]\left(\frac{\partial}{\partial t}\right)\right)$$

So long as the integration weight w is smooth in s, this expression is composed of differential operators which are smooth as maps: $S \rightarrow L(R, R^0)$, and their adjoints. It therefore defines a smooth map: $S \rightarrow L(R, R')$. As it was constructed to begin with as a composition of invertible maps, it is itself invertible, hence its inverse is smooth. However

$$M_\epsilon[s] = B[s]N_\epsilon^{-1}[s]B[s]' = \Phi[s]\tilde{B}[s]N_\epsilon^{-1}[s]\tilde{B}[s]'\Phi[s]'$$

$$= \Phi[s]((\tilde{B}[s]')^{-1}N_\epsilon(\tilde{B}[s])^{-1})^{-1}\Phi[s]'$$

whence $M_\epsilon[s]$ is smooth in s, hence so is $q_\sigma[s]$, hence so is $J_\sigma^{\rm red}[s]$.

The simulator expression we have been using for $B[s]$ is the result of high-frequency asymptotic solution of the wave equation. Systematic use of high-frequency approximation results in further simplification of M_ϵ. This is important, as the application of this operator or of N_ϵ^{-1} involves the solution of an elliptic boundary value problem. The following argument in effect allows one to replace an elliptic problem with variable coefficients by one with constant coefficients, leading to much more efficient numerical algorithms.

Note that

$$\sqrt{w}\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right) = \left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t} - \frac{1}{\sqrt{w}}\frac{\partial\sqrt{w}}{\partial z_s}\right)\sqrt{w}$$

whereas multiplication by $\sqrt{w}$ (interpreted as acting on R or R^0 as appropriate) commutes with $\partial/\partial t$. Thus

$$(\tilde{B}[s]')^{-1}N_\epsilon(\tilde{B}[s])^{-1} = \sqrt{w}$$

$$\times\left(\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right)'\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right) + \epsilon^2\left(\frac{\partial}{\partial t}\right)'\left(\frac{\partial}{\partial t}\right) + \left(\frac{1}{\sqrt{w}}\frac{\partial\sqrt{w}}{\partial z_s}\right)^2 + \frac{\partial}{\partial z_s}\left(\frac{1}{\sqrt{w}}\frac{\partial\sqrt{w}}{\partial z_s}\right)\right)$$

$$\times\sqrt{w}$$

Multiplication by $\sqrt{w}$ commutes with the translation operators T, T^0, since $w[s]$ is a function of z_s only. Thus

$$M_\epsilon[s] = \left(\phi\frac{a}{\sqrt{w}}\frac{\partial}{\partial t}\right)$$

$$\times\left(\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right)'\left(\frac{\partial}{\partial z_s} - \frac{\partial\tau}{\partial z_s}\frac{\partial}{\partial t}\right) + \epsilon^2\left(\frac{\partial}{\partial t}\right)'\left(\frac{\partial}{\partial t}\right) + \left(\frac{1}{\sqrt{w}}\frac{\partial\sqrt{w}}{\partial z_s}\right)^2 + \frac{\partial}{\partial z_s}\left(\frac{1}{\sqrt{w}}\frac{\partial\sqrt{w}}{\partial z_s}\right)\right)^{-1}$$

$$\times \left(\phi \frac{a}{\sqrt{w}} \frac{\partial}{\partial t}\right)'$$

Now suppose that 1) we choose $w = a^2$, which justifies allowing w to depend on s; 2) the data d is *oscillatory.* Then so is the solution $q_\sigma[s]$ of the dual normal equations and any of its approximations. This implies that for the price of the type of error we have already commited by adopting the geometric optics approximation in the first place we can replace the variable coefficient operator in the middle of the above expression by its principal part, hence approximate $(\tilde{B}')^{-1}N_\epsilon(\tilde{B})^{-1}$ by

$$(\tilde{B}')^{-1}N^a_\epsilon[s](\tilde{B})^{-1} \approx a[s]' \left(\left(\frac{\partial}{\partial z_s} - \frac{\partial \tau}{\partial z_s}\frac{\partial}{\partial t}\right)' \left(\frac{\partial}{\partial z_s} - \frac{\partial \tau}{\partial z_s}\frac{\partial}{\partial t}\right) + \epsilon^2 \left(\frac{\partial}{\partial t}\right)' \left(\frac{\partial}{\partial t}\right)\right) a[s]$$

$$= T'_{\tau[s]}a[s]' \left(\left(\frac{\partial}{\partial z_s}\right)' \left(\frac{\partial}{\partial z_s}\right) + \epsilon^2 \left(\frac{\partial}{\partial t}\right)' \left(\frac{\partial}{\partial t}\right)\right) a[s]T_{\tau[s]}$$

whence

$$N^a_\epsilon[s] = a[s]'N^0_\epsilon a[s]$$

where

$$N^0_\epsilon \equiv \left(\left(\frac{\partial}{\partial z_s}\right)' \left(\frac{\partial}{\partial z_s}\right) + \epsilon^2 \left(\frac{\partial}{\partial t}\right)' \left(\frac{\partial}{\partial t}\right)\right)$$

Thus finally we get more computable approximate expressions

$$B^0[s] \equiv \phi T^0_{-\tau[s]} \frac{\partial}{\partial t}$$

$$B_\epsilon[s]^* \approx a[s]^{-1}(N^0_\epsilon)^{-1}B^0[s]'$$

$$M_\epsilon[s] \approx M^0_\epsilon \equiv B^0[s](N^0_\epsilon)^{-1}B^0[s]'$$

Thus *the use of properly scaled norm in the model space has rendered the dual normal equation essentially a constant coefficient problem.* This is the analogue of the "quantitative migration" property obtained from similar norm scaling in the coherent (output least squares) formulation.

Finally it is straighforward to verify that this approximation carries through the computation of the derivative. Thus

$$\mathrm{grad}\, J^{\mathrm{red}}_\sigma[s] \approx -H_{B^0}[s]((N^0_\epsilon)^{-1}[s]B^0[s]'\tilde{q}_\sigma[s], \tilde{q}_\sigma[s])$$

where $\tilde{q}_\sigma[s]$ is an approximate solution of the approximate normal equations

$$(M^0_\epsilon + \sigma^2)q = d$$

Thus the computation of the gradient does not involve the amplitude at all. As pointed out by [6], this fact leads to the conclusion that critical points of the DSO objective function for small enough σ and noise free data are *all satisfactory solutions of the transmission inverse problem*, in the sense that these (velocities) reproduce the traveltimes of the target model within a wavelength.

Analogous formulae for the velocity gradient in the DSO formulation of reflection inversion can be derived from the Kirchhoff representation of the simulation operator (see for instance the article "Kirchhoff Simulation, Migration, and Inversion using Finite Difference Travel Times and Amplitudes" in this report, for a derivation of this representation and other references). These have the very important feature over the original (primal, unscaled) DSO/Kirchhoff velocity that the gradient of the geometric optics amplitude is not required. This is a very important practical advantage, and also morally pleasing: to define a velocity update, only traveltime information should be important.

References

[1] G. CHAVENT. Duality methods for waveform inversion. preprint, 1994.

[2] FRANÇOIS CLÉMENT and GUY CHAVENT. Waveform inversion through MBTT formulation. In E. Kleinman, T. Angell, D. Colton, F. Santosa, and I. Stakgold, editors, *Mathematical and Numerical Aspects of Wave Propagation*. SIAM, Philadelphia, 1993.

[3] G. FOLLAND. *Introduction to Partial Differential Equations*. Princeton University Press, Princeton, 1976.

[4] MARK GOCKENBACH. *An abstract analysis of differential semblance optimization*. PhD thesis, Computational and Applied Mathematics Department, Rice University, Houston, Texas, U.S.A., 1994.

[5] M. KERN and W.W. SYMES. Velocity estimation by differential semblance optimization: algorithm structure and numerical examples. *Geophysical Prospecting*, in press, 1994.

[6] HUA SONG. *On a transmission inverse problem*. PhD thesis, Computational and Applied Mathematics Department, Rice University, Houston, Texas, U.S.A., 1994.