Modification of conjugate gradient iteration to enable control of the eigenvalue range inverted

Christof Stork

ABSTRACT

The basis for chosing the scaling factors for conjugate gradient iteration is changed from the minimization of the data variance to the accurate inversion of eigenvalues over a predetermined range. This different basis enables the inversion to be controlled using knowledge of the inversion problem and produces a greater likelihood inversion while keeping the favorable characteristics of conjugate gradient intact.

INTRODUCTION

A previous paper in this report (Stork, 1988) demonstrated through comparison to Richarson's iteration how conjugate gradient fails to satisfy two desired criterion: that the inversion can be controlled, and that it produce a maximum likelihood (Aki and Richard's, 1980) inversion. A maximum likelihood inversion inverts the larger eigenvalues well even when the data does not contain significant components at that eigenvalue. Conjugate gradient, in its quest to reduce data variance in the minimum number of iterations, will not bother to invert the eigenvalues with little corresponding data energy components. However, generally we can afford to perform the very few additional iterations necessary to better invert these eigenvalues.

In reducing data variance, an additional iteration of pure conjugate gradient may invert to smaller eigenvalue rather than inverting the larger eigenvalues more accurately. In some cases, inversion to smaller eigenvalue may not be warranted. The signal to noise ratio may be too low or the desired objective may exist only above a certain eigenvalue.

In general, by being able to control the eigenvalue range inverted enables one use the modified conjugate gradient to easily analyze the characteristics of an inverse problem and then use that knowledge and other knowledge of the inverse problem in the inversion of the data. In particular, by being independent of the data components, synthetic inversion of this modified conjugate gradient can be directly compared with data applications.

It turns out that the simple modification of changing the variance function that conjugate gradient uses to determine it's scaling factors alleviates these problems. Whereas pure conjugate gradient uses variance based on data:

$$\textit{data variance} = | \, | \mathbf{b}^{(0)} - \mathbf{A}\mathbf{x}^{(m)} | \, |$$

instead we use variance based only on matching the eigenvalue inversion over a particular eigenvalue range. We will call this "eigenvalue variance."

$$\textit{eigenvalue variance} = \int_{\lambda_{min}}^{\lambda_{max}} (1.0 - \lambda \cdot \lambda^{G_n})^2 d\lambda$$

where:

λ^{G_n} = the generalized inverse achieved after n iterations.

λ_{min} & λ_{max} define the eigenvalue range over which we wish to invert.

This method can be thought of as fooling conjugate gradient into thinking that the data energy distribution is even over the desired eigenvalue range and that there is no data energy outside the range. Thus conjugate gradient will invert the eigenvalue range approximately evenly and not bother to invert the eigenvalues below the range.

IMPLEMENTING EIGENVALUE VARIANCE

For a forward problem written as:

$$\mathbf{b} = \mathbf{A}\mathbf{x},$$

The inversion achieved after the n'th iteration is represented as:

$$\mathbf{x}^{(n)} = \mathbf{A}^{G_n} \mathbf{b}^{(0)}$$

where $\mathbf{A}^{G_n}$ is the generalized inversion of $\mathbf{A}$ after n iterations.

Conjugate gradient iteration (Scales, 1987; Hestens and Stiefel, 1952) is generally written as:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \sigma_n \cdot \mathbf{p}_{(n)},$$

where:

$$\mathbf{p}_{(n)} = \mathbf{A}^T \mathbf{b}^{(n)} + \mu_n \cdot \mathbf{p}_{(n-1)}$$

$$\mu_n = \frac{|\,|\,\mathbf{A}^T \mathbf{b}^{(n)}\,|\,|}{|\,|\,\mathbf{A}^T \mathbf{b}^{(n-1)}\,|\,|}$$

σ_n is chosen to minimize the data variance, which is done by: $|\,|\,\mathbf{b}^{(0)} - \mathbf{A}\mathbf{x}^{n+1}\,|\,|$.

$$\sigma_n = (\mathbf{p}^T \mathbf{A}^T \mathbf{A}\mathbf{p})^{-1} \mathbf{p}^T \mathbf{A}^T \mathbf{b}^{(n)}$$

$\mathbf{A}^T$ = the gradient operator

$\mathbf{b}^{(n)}$ = the residual data not explained by the previous iterations, $(\mathbf{b}^{(n)} = \mathbf{b}^{(0)} - \mathbf{A}\mathbf{x}^{(n)})$.

Preconditioning and data weighting can be included in this formulation (Stork, 1988b).

Adjusting this method so the determination of the scaling factor, σ_n, is performed using the eigenvalue variance requires the reformulation of this iteration procedure in terms of the eigenvalues. The effect of each iteration on the eigenvalue inversion can be determined by substituting the singular value decomposition for $\mathbf{A}$, $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$:

$$\mathbf{V}^T \mathbf{x}^{(n+1)} = \mathbf{V}^T \mathbf{x}^{(n)} + \sigma_n \cdot \mathbf{V}^T \mathbf{p}_{(n)},$$

$$\mathbf{V}^T \mathbf{p}_{(n)} = \Sigma \mathbf{U}^T \mathbf{b}^{(n)} + \mu_n \cdot \mathbf{V}^T \mathbf{p}_{(n-1)}$$

$$\mu_n = \frac{|\,|\,\mathbf{V}^T \Sigma \mathbf{U}^T \mathbf{b}^{(n)}\,|\,|}{|\,|\,\mathbf{V}^T \Sigma \mathbf{U}^T \mathbf{b}^{(n-1)}\,|\,|} = \frac{|\,|\,\Sigma \mathbf{U}^T \mathbf{b}^{(n)}\,|\,|}{|\,|\,\Sigma \mathbf{U}^T \mathbf{b}^{(n-1)}\,|\,|}$$

$$\sigma_n = (\mathbf{p}^T \mathbf{V}^T \Sigma^2 \mathbf{V}^T \mathbf{p})^{-1} \mathbf{p}^T \mathbf{V}^T \Sigma \mathbf{U} \mathbf{b}^{(n)}$$

$$\mathbf{U}^T \mathbf{b}^{(n)} = \mathbf{U}^T \mathbf{b}^{(0)} - \Sigma \mathbf{V}^T \mathbf{x}^{(n)}$$

$$\mathbf{V}^T \mathbf{x}^n = \Sigma^{G_n} \mathbf{U}^T \mathbf{b}^{(0)}$$

The $\mathbf{U}^T \mathbf{b}$ and $\mathbf{V}^T \mathbf{x}$ operation separate the data and model into their eigenvector components. Since Σ is diagonal, these matrix equations can be rewritten as scalar

equations for each eigenvalue, λ, and its data and model components. We will treat the terms of this equation as continuous functions of the eigenvalues with the following substitutions:

$$\mathbf{U}^T \mathbf{b} \rightarrow U(\lambda)$$
$$\mathbf{V}^T \mathbf{x} \rightarrow V(\lambda)$$
$$\mathbf{V}^T \mathbf{p} \rightarrow P(\lambda)$$
$$\Sigma \rightarrow \lambda$$

where $U(\lambda)$, $V(\lambda)$, & $P(\lambda)$ are defined as:

$$U^n(\lambda) \cdot d\lambda = \lim_{d\lambda \rightarrow 0} \int_{\lambda}^{\lambda + d\lambda} \mathbf{b}^{(n)^T} \cdot \mathbf{u}(\lambda')\, d\lambda'$$

$$V^n(\lambda) \cdot d\lambda = \lim_{d\lambda \rightarrow 0} \int_{\lambda}^{\lambda + d\lambda} \mathbf{x}^{(n)^T} \cdot \mathbf{v}(\lambda')\, d\lambda'$$

$$P^n(\lambda) \cdot d\lambda = \lim_{d\lambda \rightarrow 0} \int_{\lambda}^{\lambda + d\lambda} \mathbf{p}^{(n)^T} \cdot \mathbf{v}(\lambda')\, d\lambda'$$

and

$\mathbf{u}(\lambda')$ = the data space eigenvector at eigenvalue λ' .

$\mathbf{v}(\lambda')$ = the model space eigenvector at eigenvalue λ' .

These functions represent the distribution of energy in the data and model over the eigenvalue range. Note that the distribution of eigenvalues is included in this function.

The new equations are:

$$V^{n+1}(\lambda) = V^n(\lambda) + \sigma_n \cdot P^n(\lambda)$$

$$P^n(\lambda) = \lambda U^n(\lambda) + \mu_n \cdot P^{n-1}(\lambda)$$

$$\mu_n = = \frac{\int_0^\infty \lambda^2 \left(U^n(\lambda) \right)^2 d\lambda}{\int_0^\infty \lambda^2 \left(U^{n-1}(\lambda) \right)^2 d\lambda}$$

$$\sigma_n = \frac{\int_0^\infty P^n(\lambda) \lambda U^n(\lambda) d\lambda}{\int_0^\infty P^n(\lambda)^2 \cdot \lambda^2 d\lambda}$$

$$U^n(\lambda) = U^0(\lambda) - \lambda \cdot V^n(\lambda)$$

$$V^n(\lambda) = \lambda^{G_n} U^0(\lambda)$$

We seek a formula for $Q^n(\lambda) = \lambda \cdot \lambda^{G_n}$ after each iteration which represents the inversion level of the eigenvalues. (λ^{G_n} represents the generalized inversion of eigenvalue λ achieved after n iterations.) We rewrite the equation in terms of $Q^n(\lambda) = \lambda \cdot \lambda^{G_n}$ and substitute $U^0(\lambda) \cdot \tilde{P}^n(\lambda) = P^n(\lambda) \cdot \lambda$ to simplify the equations.

$$U^n(\lambda) = \left(1.0 - Q^n(\lambda) \right) \cdot U^0(\lambda)$$

$$\tilde{P}^n(\lambda) = \lambda^2 \cdot \left(1.0 - Q^n(\lambda) \right) + \mu_n \cdot \tilde{P}^{n-1}(\lambda)$$

$$\mu_n = \frac{\int_0^\infty \lambda^2 \cdot \left(1.0 - Q^n(\lambda) \right)^2 \cdot U^0(\lambda)^2 d\lambda}{\int_0^\infty \lambda^2 \cdot \left(1.0 - Q^{n-1}(\lambda) \right)^2 \cdot U^0(\lambda)^2 d\lambda}$$

$$Q^{n+1}(\lambda) = Q^n(\lambda) + \sigma_n \cdot \tilde{P}^n(\lambda)$$

$$\sigma_n = \frac{\int_0^\infty \tilde{P}^n(\lambda) \cdot \Big(1.0 - Q^n(\lambda)\Big) \cdot U^0(\lambda)^2 d\lambda}{\int_0^\infty \tilde{P}^n(\lambda)^2 \cdot U^0(\lambda)^2 d\lambda}$$

Normally, the function $U^0(\lambda)$ represents the distribution of energy in the data over the eigenvalue range. The data variance that conjugate gradient tries to minimize,

$$data\ \ variance\ =\ \|\ \mathbf{b}^{(0)} - \mathbf{A}\mathbf{x}^{(m)}\ \|$$

can be rewritten using our continuous functions (Stork, 1988b) as:

$$\int_0^\infty U^0(\lambda)^2 \cdot \Big(1 - Q(\lambda)\Big)^2 d\lambda$$

Our complaint with conjugate gradient is that it may invert eigenvalues below a desired level and that it may invert a larger eigenvalue range poorly when it contains few data energy components. However, with the equations above, we can redefine $U^0(\lambda)$ to suite our own desires. If we desire the inversion to not invert the eigenvalues below a certain point, we set this function to zero below that eigenvalue. To invert eigenvalues approximately evenly above that point, the function is set to a constant above the eigenvalue. Thus, the proposed value of $U^0(\lambda)$ to invert over an eigenvalue range $\lambda_{\min} < \lambda < \lambda_{\max}$ is:

$$U^0(\lambda) = 1.0 \quad \text{for } \lambda_{\min} < \lambda < \lambda_{\max}$$
$$U^0(\lambda) = 0.0 \quad \text{for } \lambda < \lambda_{\min}$$

If, after several iterations, we decide to invert to lower eigenvalue than originally specified, we need only change the function, $U^0(\lambda)$ to include the greater eigenvalue range, and continue the iterations.

NUMERICAL EXAMPLE

Figure 2 is an example of the function $\lambda \cdot \lambda^G$ produced using this choice for $U^0(\lambda)$, which is shown in Figure 1. The corresponding function for $\lambda \cdot \lambda^G$ produced

using Richarson's iteration with Chebyshev acceleration factors (Stork, 1988) is also plotted for comparison. While Chebyshev iteration has uniform error over the specified eigenvalue range, conjugate gradient has inverted the larger eigenvalues more accurately than the lower ones in the range. The difference is not very significant, but it is enough for the "eigenvalue variance," as defined earlier, to be about half for conjugate gradient as it is for Chebyshev iteration.

However, a main advantage of conjugate gradient iteration over Chebyshev iteration is that the number of iterations need not be specified before hand. Conjugate gradient iteration can be continued until the desired accuracy over the eigenvalue range is achieved. It can be easily restarted when inversion to smaller eigenvalue is desired. Chebyshev iteration, however, would require the number of iterations and eigenvalue range be specified at the start. Changing either requires restarting from beginning.

CONCLUSIONS

The use of "eigenvalue variance" instead of the standard data variance for the conjugate gradient iteration achieves a result closer to the maximum likelihood inversion and the eigenvalue range inverted can be easily controlled. The speed of conjugate gradient and its ability to arbitrarily stop and continue the inversion are kept intact.

ACKNOWLEDGEMENTS

This work is a direct continuation of the my thesis research performed under Rob Clayton at the Caltech Seismology Lab.

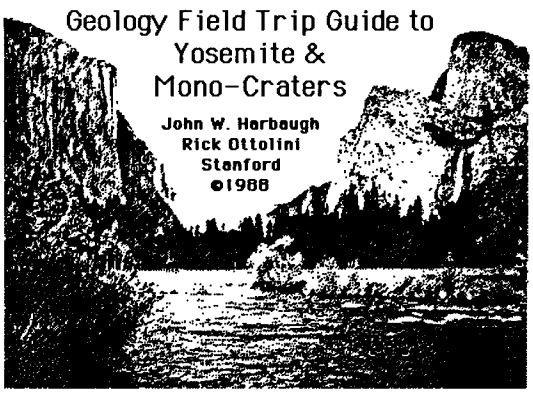
My interest and motivation for this work comes almost entirely from discussions with John Scales of Amoco.

REFERENCES

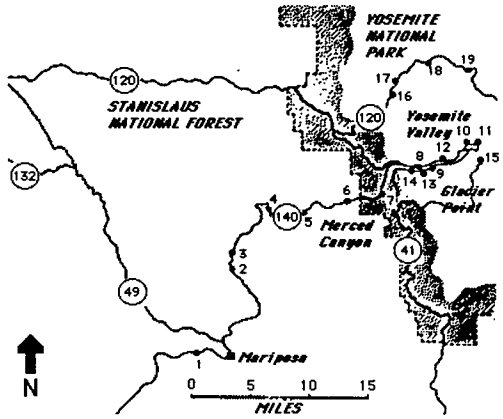
Hestens, M. and Stiefel, E., 1952, Methods of conjugate gradients for solving linear systems: Nat. Bur. Standards J. Res., **49**, 409-436.

Scales, J., 1987, Tomographic inversion via the conjugate gradient method: Geophysics, **52**, 179-185.

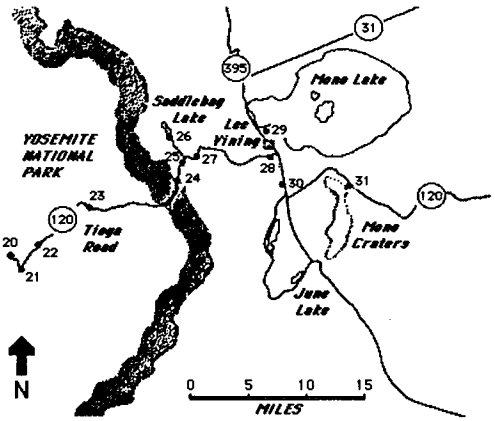
Stork, C., 1988, Comparison of Richarson's iteration with Chebyshev acceleration factors to conjugate gradient iteration: SEP–**57**.



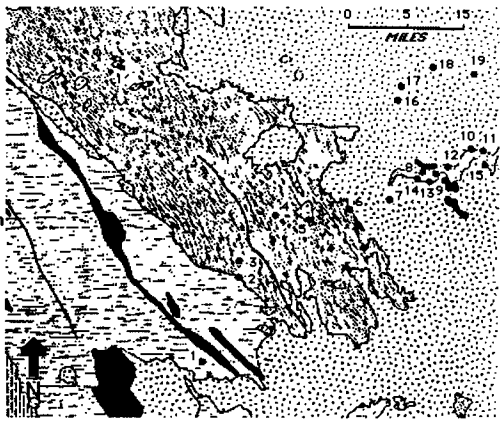
About the field guide...
About Yosemite Mono Craters...
Geology map
Road map B
California map



About the field guide...
About Yosemite Mono Craters...
Geology map
Road map A
California map



About the field guide...
About Yosemite Mono Craters...
Geology map B
Road map
Geologic Column



About the field guide...
About Yosemite Mono Craters...
Geology map A
Road map
Geologic Column