Migration preconditioning with Curvelets

Peyman P. Moghaddam* and Felix J. Herrmann
University of British Columbia, Canada

Summary

In this paper, the property of Curvelet transforms for preconditioning the migration and normal operators is investigated. These operators belong to the class of Fourier integral operators and pseudo-differential operators, respectively. The effect of this preconditioner is shown in term of improvement of sparsity, convergence rate, number of iteration for the Krylov-subspace solver and clustering of singular(eigen) values. The migration operator, which we employed in this work is the common-offset Kirchhoff-Born migration.

Introduction

The Curvelet transform introduced by Candés[2002] has so far primarily been utilized for image processing and medical imaging. The ability of this non-separable Wavelet transform to preserve the edges in a noisy image is proven to be superior than the performance of the separable 2-D Wavelet transform.

Curvelets are multi-scale, multi-directional, anisotropic, local both in space and in Fourier domain and almost orthogonal, (i.e., they are tight frames). Candés and Demanet[2003] explore another interesting property of Curvelets, which makes quite useful for exploration seismology. They proved that a Curvelet remains Curvelet-like (i.e., similar to another curvelet) under the action of Fourier Integral Operators (FIO). This property may allow us to introduce a fast algorithm for the migration operator, which is a class of FIO.

To investigate this property numerically, we precondition the migration operator with the Curvelet transform. Subsequently, we analyze the properties of the preconditioned and original operators. We also investigate the preconditioning property of the Curvelet transform for the normal operator (i.e., migration/demigration operator), which is mathematically a pseudo-differential operator. We investigate (i) the sparsity; (ii) the number of iterations for the conjugate-gradient based solver; (iii) the convergence rate and singular-value clustering for the operator after preconditioning and the operator itself both for migration and normal operator.

A common offset Kirchhoff-Born migration/scattering operator with constant velocity background is utilized for this investigation.

Problem Formulation

In general form[Rickett 2003], the linear forward scatter-

ing problem can be written:

$$Km = d \quad (1)$$

Where d is data, K is the scattering operator and m the model parameters. The solution to equation (1) is of the following form:

$$\hat{m}_{\ell 2} = (K^*K)^{-1}K^*d, \quad (2)$$

where K^* refers to the adjoint (or Hermitian) operator of K .

Both K and K^* are Fourier Integral Operator (FIO) and represent the scattering and migration, respectively. The product (K^*K) is defined as the normal/Hessian/Gramm operator and belongs to the class of operators known as pseudo-differential operators. The computationally expensive part in equation (2) is to invert for the normal operator. This is practically impossible (except for explicit closed-form solutions for the inverse of the normal operator) and the solution for equation (1) is achieved through a limited number of iterations for the Krylov-subspace based method[Saad, 2003].

The success of Krylov-subspace based methods depends on two main characteristics of the operator: first its sparsity and second the clustering of its singular values. The sparsity means the number of nonzero elements in the matrix representing the operator. Thus, the sparser the operator the fewer calculations are needed to compute matrix-vector products and hence faster computing time and less memory requirements. Clustering of singular values means concentration of almost all of the singular values in a few specific locations on the real axis far away from zero. The denser the clouds of singular values (i.e., more clustered) the more accurate each iteration converges to the exact solution (i.e., faster convergence).

Rickett[2003] proposed using a preconditioning matrix for K which results to faster convergence and robust recovery of model parameters. He came to the question: what is a good choice for preconditioning matrix (operator)? Candés and Demanet showed that a Curvelet remain Curvelet-like under a FIO. Since migration operator is a FIO, we introduce the following preconditioned system of equations:

$$CKC^{-1}Cm = Cd \implies \tilde{K}\tilde{m} = \tilde{d} \quad (3)$$

in which $\tilde{K} = CKC^{-1}$ is the preconditioned K , $\tilde{m} = Cm$ the Curvelet transform of m and $\tilde{d} = Cd$ the Curvelet transform of d . Throughout this paper, C stands for the

Migration preconditioning with Curvelets

Curvelet transform, C^{-1} for the inverse Curvelet transform and C^T is the adjoint (or transpose) of the transform.

Now our aim is to show that $\tilde{K}$ is improved in its main matrix properties compared to K . We compare sparsity and clustering of singular values of K . In addition, we study the number of iterations for Krylov-subspace based methods for solution of the preconditioned system (equation (3)) compared to the original system (equation (1)).

In some applications, instead of using a conjugate gradient based method to solve equation (1), an easy-to-invert approximation of the normal operator (i.e., K^*K) is used and plugged into equation (2). For example, Claerbout [1994] used the diagonal elements of the normal operator applied on the reference model and replaced the action of the normal operator by this diagonal weighting. Preconditioning of the normal operator will also be useful if the normal operator can be better diagonalized (i.e., sparser). We applied this property of preconditioning to replace the preconditioned normal operator with its approximation [see also other contributions by authors to the Proceedings of this Conference].

In this work, we also investigate the effect of preconditioning using the Curvelet transform on the normal operator. In this regard, we define a new system of equations, which is based on the normal operator and weighted normal operator as follows:

$$\psi = K^*K \implies \tilde{\psi} = CK^*KC^{-1} \quad (4)$$

in which $\tilde{\psi}$ stands for weighted ψ .

The sparsity of a matrix is defined as number of nonzero elements divided by the total number of elements in the matrix. Since K and $\tilde{K}$ (and obviously ψ and $\tilde{\psi}$) are extremely large-sized matrices and defined in the form of operators (matrix-vector products), measuring the exact sparsity is computationally prohibitive. Nevertheless, sampling a few columns of these operators can result to an approximation of the sparsity. Here is the strategy. Suppose we want to estimate the sparsity of a matrix $A_{M \times N}$, we randomly sample a column of A by simply multiplying it by a vector e_i with random $i \in [1, 2, \dots, N]$. e_i is an all-zero vector, except it has a single one in its i th location. An approximation of sparsity is achieved using the following formula:

$$\hat{s}_A = \frac{\sum_{i=1}^T M - Nz_i}{TM} \times 100(\text{percentage}), \quad (5)$$

where T is the number of realizations, Nz_i the number of nonzero elements in each realization.

In addition to sparsity, clustering is also an important property of the matrix structure. To measure how the singular values of the matrix A are clustered, we need to compute the singular values of A . The method of Lanczos bidiagonalization [Golub 1996] can be utilized to decompose a non-square matrix A into the following form after

k iterations:

$$A = U_{k+1}B_kV_k^T. \quad (6)$$

The matrices U_{k+1} and V_k are orthogonal matrices and k is the number of iterations in the Lanczos method [Golub 1996], B_k is a lower bidiagonal matrix:

$$B_k = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ \beta_2 & \alpha_2 & \cdots & 0 \\ 0 & \beta_3 & \ddots & \vdots \\ 0 & \cdots & \ddots & \alpha_k \\ 0 & 0 & \cdots & \beta_{k+1} \end{pmatrix} \quad (7)$$

It has been shown [Golub 1996] that the singular values of B_k are an approximation to singular values of A . A disadvantage of Lanczos method is that the estimated singular values in this method do not belong to the most dominant singular values of A . In other words, the estimated singular values of A using Lanczos method are randomly chosen among all singular values. This makes the process of clustering difficult since we are looking for a distance between adjacent singular values.

Larsen [1998] proposed a robust and efficient Lanczos bidiagonalization with partial reorthogonalization (LBDPRO) method for estimating the most dominant adjacent singular values of a matrix A . Since the migration operator K and the preconditioned migration operator $\tilde{K}$ are both non-square matrices, we use LBDPRO method to estimate the singular values of them. Larsen [1998] proposed also the method of Lanczos tridiagonalization with partial reorthogonalization (LTDPRO), when the matrix A is square and symmetric positive definite. In this method A is decomposed into a new form of the matrix product:

$$A = Q_k T_k Q_k^T \quad (8)$$

in which Q_k is an orthogonal matrix and T_k is a tridiagonal matrix that reads:

$$T_k = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & \cdots & 0 \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & \ddots & \ddots & \beta_{k-1} \\ 0 & \cdots & & \beta_{k-1} & \alpha_k \end{pmatrix}. \quad (9)$$

The Eigen values of T_k are Ritz values for the matrix A and these are good approximations for its eigen values.

The normal operator (ψ as in equation (4)) is symmetric positive definite and since the Curvelet transform is a tight frame with unit energy (i.e., $|Cx|_2 = |x|_2, \forall x \in R^n$), we can reliably assume that it is close to orthonormal so that $C \approx C^{-T}$ and $C^T \approx C^{-1}$. Under this assumption, the preconditioned normal operator ($\tilde{\psi}$ as in equation (4))

Migration preconditioning with Curvelets

is symmetric positive definite too. Therefore, we can utilize the LTDPRO method to estimate the eigen-values of the normal operator and the preconditioned normal operator.

A disadvantage of Partial ReOrthogonalization(PRO) methods is that they need to reorthogonalize the matrices U_k , V_k (for LBDPRO) and Q_k (for LTDPRO) for each iteration k . Since these matrices are quite large, this operation consumes a lot of memory and limits our ability to estimate more than a few tens of singular values. Nevertheless, since these values belong to the most dominant singular values, they are good measure for comparison.

We also investigated the difference between convergence rate of preconditioned and original operators. We show that residual error decays faster for the preconditioned operator compared to the operator itself. In this regard, we solve for two different systems of equations, first for the migration and then for the preconditioned migration, i.e. for

$$Kx = y \xRightarrow{\text{Preconditioned}} \tilde{K}\tilde{x} = \tilde{y} \quad (10)$$

and for the preconditioned normal operator

$$\psi x = y \xRightarrow{\text{Preconditioned}} \tilde{\psi}\tilde{x} = \tilde{y} \quad (11)$$

Since K and $\tilde{K}$ are not square, we use GMRES method in equation (10) and (11) to solve for the two systems of equations. Since ψ and $\tilde{\psi}$ are symmetric positive definite, we use the Conjugate-Gradient method. We plot the residual error of these two systems of equations versus iteration number and compare them.

Results

Table (1) shows the sparsity measurements and the comparisons between K and $\tilde{K}$, ψ and $\tilde{\psi}$ for common-offset the Born-Kirchhof migration operator with constant background velocity. It can be seen from this table that preconditioning significantly improves the sparsity for both migration and normal operators. This verifies the claim of Candés and Demanet(2003) that Curvelets remain curvelet-like under the action of the migration and normal operators. The second row in this table shows the number of iterations, needed to reduce the residual error to a value smaller than a specific tolerance.

Figure 1 shows the evaluation of the singular value clustering for K and $\tilde{K}$. As it can be seen in this figure, singular values of $\tilde{K}$ are concentrating in and around the abscissa -5 , while the singular values for K tempt to concentrate around zero. Figure 2 shows the evaluation for the eigen-values clustering for ψ and $\tilde{\psi}$. We can see similar behavior of the singular values in this figure as well. These figures show that the singular(eigen) values are more clustered for preconditioned operators compared to the unconditioned operators themselves.

To measure the convergence rate, we need to compute the residual error. Residual error is defined as

$$r_k = \frac{|y - Ax_k|_2}{|y|_2} \xRightarrow{\text{Preconditioned}} \tilde{r}_k = \frac{|\tilde{y} - \tilde{A}\tilde{x}_k|_2}{|\tilde{y}|_2}. \quad (12)$$

A in the above formula can be K , according to equation(10), ψ according to equation(11). For both equations (10) and (11), we apply the above residual error to compare original and preconditioned systems.

Figure 3, shows the residual error versus the iteration number for the GMRES method used to solve the system in equation (10). Figure 4, shows the residual error versus the iteration number for the Conjugate-Gradient method for the solution of equation (11). As it can clearly be seen, preconditioning improves the convergence rate.

Operator	K	$\tilde{K}$	ψ	$\tilde{\psi}$
Sparsity	52.5	78.2	62.3	91.7
Iteration	> 100	44	58	23

Table 1: Sparsity and number of iterations for the solution of equations (10) and (11). First row is sparsity as defined in equation (5). Second row is the number of iterations needed to be taken by the GMRES-method (for K and $\tilde{K}$) and by the CG method (for ψ and $\tilde{\psi}$) for the residual error to be smaller than 10^{-3} .

Conclusions

In this paper, the effects of the Curvelet transform for preconditioning of migration and normal operator were investigated. This preconditioning can:

- improve the sparseness of the operator; The sparsity of both migration and normal operator is increased after preconditioning.
- improve the clustering of singular(eigen) values of the operator; For the preconditioned operators, the singular values shifted away from zero and have a tendency to concentrate in a point compared to the clustering of the operator itself.
- improve the convergence rate of the operator; The convergence rate for preconditioned normal operators is surprisingly faster than for the normal operator itself.

We demonstrated that by employing the Curvelet transform as preconditioner, we were able to enhance certain important matrix characteristics for the migration and normal operators. The results for this preconditioning can be exploited to introduce faster and robust solutions for existing migration operators.

Acknowledgements

The authors thank to Emmanuel Candés for making

Migration preconditioning with Curvelets

their Curvelet transform code available. We also would like to thank Mauricio Sacchi for his migration code. We are also grateful to Rasmus Munk Larsen and Chen Grief to make us available the PROPACK software. This work was in part financially supported by a NSERC Discovery Grant.

Fig. 1: Singular Value Trends for migration and preconditioned migration.

Fig. 2: Singular value trends for the normal and preconditioned normal operators.

References

- [1] E. Candès and F. Guo, New multiscale transforms, minimum total variation synthesis: Application to edge preserving image reconstruction. *Signal Processing*, pages 1519-1543, 2002
- [2] E. Candès and L. Demanet, Curvelets and Fourier Integral Operators ,C. R. Acad. Sci. Paris, Ser. I 336 (2003) 395-398
- [3] J.E. Rickett, Illumination-based normalization for wave-equation depth migration, *Geophysics*, vol.68, No.4, P.1371-1379, 2003
- [4] Y. Saad, Iterative methods for sparse linear systems, *SIAM*, 2003

[5] J. Claerbout and D. Nichols, 1994, Spectral Preconditioning: *Stanford Exploration Project Report*, 82, 183-186

[6] G.H. Golub and C.F. Van Loan, Matrix Computations, *The John Hopkins University Press*, Third edition, 1996

[7] R.M. Larsen, Lanczos bidiagonalization with partial reorthogonalization, Department of computer science, Aarhus University, *Technical Report*, DAIMI PB-357, 1998

Fig. 3: Residual error of GMRES method versus the number of iterations needed for the solution of equation (10)

Fig. 4: Residual error of Conjugate-Gradient method versus the number of iterations needed for the solution of equation (11)