Singular Vector Analysis for Atmospheric Chemical Transport Models

WENYUAN LIAO AND ADRIAN SANDU
Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

GREGORY R. CARMICHAEL AND TIANFENG CHAI
Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa

(Manuscript received 14 March 2005, in final form 27 October 2005)

ABSTRACT

The singular vectors of a chemical transport model are the directions of maximum perturbation growth over a finite time interval. They have proved useful for the estimation of error growth, the initialization of ensemble forecasts, and the optimal placement of adaptive observations. The aim of this paper is to address computational aspects of singular vector analysis for atmospheric chemical transport models. The distinguishing feature of these models is the presence of stiff chemical interactions. A projection approach to preserve the symmetry of the tangent linear–adjoint operator for stiff systems is discussed, and extended to 3D chemical transport simulations. Numerical results are presented for a simulation of atmospheric pollution in East Asia in March 2001. The singular values and the structure of the singular vectors depend on the length of the simulation interval, the meteorological data, the location of the optimization region and the selection of optimization species, the choice of error norms, and the size of the optimization region.

1. Introduction

The singular vectors of a chemical transport model are the directions of fastest perturbation growth over a finite time interval. Singular vector analysis was introduced in meteorology by Lorenz (1965), who computed the largest error growth rates in an idealized model of the atmosphere. The adjoint technique was used by Molteni and Palmer (1993) and Mureau et al. (1993) to compute singular vectors for meteorological models. Singular vector analysis of general circulation models with millions of variables is now possible (see e.g., Buizza and Palmer 1995; Li et al. 2005).

Applications of singular vector analysis in numerical weather prediction include the following: 1) normal mode analysis of atmospheric flow instability, estimation of error growth, and the assessment of atmospheric predictability (Borges and Hartmann 1992; Ehrendorfer and Tribbia 1997; Molteni and Palmer 1993; Oortwijn 1998); 2) initialization of ensemble forecasts (Molteni et al. 1996; Mureau et al. 1993); and 3) estimation of the optimal placement of adaptive observations (Buizza and Montani 1999; Gelaro et al. 1998; Lorenz and Emanuel 1998; Palmer et al. 1998).

Numerous studies have shown that the structure of singular vectors in atmospheric general circulation models is determined by the following: 1) the atmospheric episode under consideration (Farrell 1988; Barkmeijer et al. 2001); 2) model physics (e.g., the treatment of boundary layer processes; Barkmeijer et al. 2001; Buizza and Montani 1999; Buizza and Palmer 1995; Ehrendorfer et al. 1999; Mahfouf 1999); 3) model resolution (Buizza and Montani 1999; Buizza and Palmer 1995; Ehrendorfer et al. 1999; Mahfouf 1999); and 4) the particular choice of error norms (Kuang 2004).

The objective of this work is to study the singular vectors for atmospheric chemical transport models. The distinguishing feature of these models is the presence of chemical interactions between tracer species, which leads to a very stiff, nonlinear system of partial differential equations. This poses nontrivial challenges in the computation of singular vectors, and allows interesting interpretations stemming from the complex interactions between emission sources, chemical transformations, transport, and deposition processes.
While in the study of atmospheric dynamics the dominant singular vectors are associated with unstable modes, in the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain (e.g., where the model prediction needs to be improved). In this paper we illustrate the use of singular vectors to describe uncertainties in the initial conditions. Other very important sources of uncertainty in air quality models that need to be quantified in real applications are the emissions, the meteorological fields, the deposition velocities, and the top and lateral boundary conditions for regional models.

The paper is organized as follows. In section 2 we introduce the chemical transport model singular vectors and discuss two of their possible applications in the context of air pollution modeling. Computational aspects of chemical singular vectors are discussed in section 3. An introduction to chemical transport modeling and the formulation of the tangent linear and adjoint models is presented in section 4. Several possible perturbation norms are discussed in section 5. Numerical results from a simulation of air pollution in East Asia are shown in section 6. Section 7 summarizes the main findings of this work.

2. Singular vectors and chemical transport models

An atmospheric chemical transport model (CTM) propagates the model state forward in time from the initial state \( \mathbf{x}(t_0) \) to the final state \( \mathbf{x}(t_F) \). With \( \mathcal{M} \) denoting the model solution operator we have

\[
\mathbf{x}(t_F) = \mathcal{M}_{t_0 \rightarrow t_F} \mathbf{x}(t_0). \tag{1}
\]

Perturbations (small errors) evolve according to the tangent linear model (TLM)

\[
\delta \mathbf{x}(t_F) = \mathbf{M}_{t_0 \rightarrow t_F} \delta \mathbf{x}(t_0), \tag{2}
\]

and adjoint variables according to the adjoint model

\[
\lambda(t_F) = \mathbf{M}^*_{t_F \rightarrow t_0} \lambda(t_F). \tag{3}
\]

Here \( \mathbf{M} \) and \( \mathbf{M}^* \) denote the solution operators of the two linearized models. A detailed description of chemical transport models, and the corresponding tangent linear and adjoint models, will be given in section 4. The error covariance matrix propagates from \( \mathbf{P}(t_0) \) to \( \mathbf{P}(t_F) \) according to

\[
\mathbf{P}(t_F) = \mathbf{M}_{t_0 \rightarrow t_F} \mathbf{P}(t_0) \mathbf{M}^*_{t_F \rightarrow t_0} + \mathbf{Q}. \tag{4}
\]

The additional term \( \mathbf{Q} \) represents the covariance of the model errors.

\( a. \) Singular vectors

Singular vectors determine the most rapidly growing perturbations in the atmosphere. The magnitude of the perturbation at the initial time \( t_0 \) is measured in the \( L^2 \) norm defined by a symmetric positive definite matrix \( \mathbf{A} \)

\[
\| \delta \mathbf{x}(t_0) \|_{\mathbf{A}}^2 = \langle \delta \mathbf{x}(t_0), \mathbf{A} \delta \mathbf{x}(t_0) \rangle, \tag{5}
\]

where \( \langle \cdot, \cdot \rangle \) denotes the scalar product in \( L^2 \). Similarly, the perturbation magnitude at the final time \( t_F \) is measured in a seminorm defined by a semipositive definite matrix \( \mathbf{B} \)

\[
\| \delta \mathbf{x}(t_F) \|_{\mathbf{B}}^2 = \langle \delta \mathbf{x}(t_F), \mathbf{B} \delta \mathbf{x}(t_F) \rangle. \tag{6}
\]

We call the norms (5) and (6) squared the “perturbation magnitudes.” The ratio between perturbation magnitudes at \( t_F \) and \( t_0 \) offers a measure of error growth:

\[
\sigma^2 = \frac{\| \delta \mathbf{x}(t_F) \|_{\mathbf{B}}^2}{\| \delta \mathbf{x}(t_0) \|_{\mathbf{A}}^2} = \frac{\langle \delta \mathbf{x}(t_F), \mathbf{B} \delta \mathbf{x}(t_F) \rangle}{\langle \delta \mathbf{x}(t_0), \mathbf{A} \delta \mathbf{x}(t_0) \rangle} = \frac{\langle \delta \mathbf{x}(t_0), \mathbf{M}_{t_0 \rightarrow t_F} \mathbf{B} \mathbf{M}_{t_0 \rightarrow t_F} \delta \mathbf{x}(t_0) \rangle}{\langle \delta \mathbf{x}(t_0), \mathbf{A} \delta \mathbf{x}(t_0) \rangle} = \frac{\langle \delta \mathbf{x}(t_0), \mathbf{M}^*_{t_F \rightarrow t_0} \mathbf{B} \mathbf{M}_{t_0 \rightarrow t_F} \delta \mathbf{x}(t_0) \rangle}{\langle \delta \mathbf{x}(t_0), \mathbf{A} \delta \mathbf{x}(t_0) \rangle}. \tag{7}
\]

In (7) we use the fact that perturbations evolve in time according to the dynamics of the tangent linear model in (2).

Singular vectors are defined as the directions of maximal error growth, that is, the vectors \( \mathbf{s}_k(t_0) \) that maximize the ratio \( \sigma^2 \) in (7). These directions are the solutions of the generalized eigenvalue problem:

\[
\mathbf{M}^*_{t_F \rightarrow t_0} \mathbf{B} \mathbf{M}_{t_0 \rightarrow t_F} \mathbf{s}_k(t_0) = \sigma_k^2 \mathbf{A} \mathbf{s}_k(t_0). \tag{8}
\]

The left side of (8) involves one integration with the tangent linear model followed by one integration with the adjoint model. The eigenvalue problem (8) can be solved efficiently using the software package ARPACK (available online at http://www.caam.rice.edu/software/ARPACK) or its parallel version PARPACK (available online at http://www.caam.rice.edu/~kristyn/parpack_home.html).

Using the square root of \( \mathbf{A} \), the generalized eigenvalue problem (8) can be reduced to an ordinary eigenvalue problem:

\[
\mathbf{A}^{-1/2} \mathbf{M}^*_{t_F \rightarrow t_0} \mathbf{B} \mathbf{M}_{t_0 \rightarrow t_F} \mathbf{A}^{-1/2} \mathbf{v}_k(t_0) = \sigma_k^2 \mathbf{v}_k(t_0),
\]

where \( \mathbf{v}_k(t_0) = \mathbf{A}^{1/2} \mathbf{s}_k(t_0) \). \tag{9}

Furthermore, \( \mathbf{v}_k(t_0) \) are the left singular vectors in the singular value decomposition:

\[
\mathbf{B}^{1/2} \mathbf{M}^*_{t_F \rightarrow t_0} \mathbf{A}^{-1/2} = \mathbf{U} \text{ diag}(\sigma_k) \mathbf{V}^T,
\]

where \( \sigma_k \mathbf{u}_k(t_F) = \mathbf{B}^{1/2} \mathbf{s}_k(t_F) \). \tag{10}
The singular vectors $s_k$ are $A$ orthogonal at $t_0$ and $B$ orthogonal at $t_F$:

$$\langle s_k(t_0), A s_j(t_0) \rangle = 0 \quad \text{and} \quad \langle s_k(t_F), B s_j(t_F) \rangle = 0$$

for $j \neq k$. (11)

The singular value decomposition of the linear operator $M_{t_F \rightarrow t_0}$, with the $A$ scalar product at $t_0$ and the $B$ scalar product at $t_F$, has the left singular vectors $s_k(t_0)$ and the right singular vectors $s_k(t_F)$. If the same norms are used at the initial and at the final times the singular values $\sigma_k$ can be interpreted as the error amplification factors along each direction $s_k$.

A special set of perturbation norms is provided by the choice $B = I$ and $A = P(t_0)^{-1}$. In this case the resulting singular vectors $s_k(t_F)$ evolve into the leading eigenvectors $P(t_F) s_k(t_F)$ of the forecast error covariance matrix $P(t_F)$:

$$P(t_F) s_k(t_F) = \sigma_k^2 s_k(t_F).$$

(12)

The eigenvectors $s_k(t_F)$ are called analysis error covariance singular vectors (Ehrendorfer and Tribbia 1997; Barkmeijer et al. 1999). Since the leading eigenvectors of $P(t_F)$ are the directions of maximum variance of forecast error, the singular vectors define the directions along which we must do a good job of analysis in order to minimize the forecast error at $t_F$. We assume that the model error in (4) is negligible over the period $[t_0, t_F]$. From (12) it follows that the singular vectors are the solutions of the following generalized eigenvalue problem:

$$M_{t_F \rightarrow t_0}^* M_{t_0 \rightarrow t_F} s_k(t_0) = \lambda_k P(t_0)^{-1} s_k(t_0) = \lambda_k (\nabla J) s_k(t_0).$$

(13)

When calculated as above the singular vectors are referred to as Hessian singular vectors (HSV; Barkmeijer et al. 1998). The second relation comes from the fact that the Hessian matrix of the analysis cost function $J$ in the variational analysis system is an estimate of the inverse of the analysis covariance matrix. This motivates the name Hessian singular vectors for the solutions $s_k(t_0)$ of the eigenproblem in (13).

In the study of atmospheric dynamics the dominant singular vectors are associated with unstable modes. In the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain (e.g., where the model prediction needs to be improved). Limited subdomain studies are also of interest in numerical weather prediction (Hersbach et al. 2003).

b. Initialization of ensemble forecasts

A critical element for accurate simulations is the use of observational data to constrain model predictions. Widely used data assimilation techniques include three-dimensional variational (3DVAR), four-dimensional variational (4DVAR), Kalman filter, and ensemble nonlinear filters. Kalman filter techniques are discussed in Daley (1991) and Jazwinski (1970). Consider a set of observations $y$ at $t_F$ (assumed, for simplicity, to be a linear function of model state, $y = Hx$). The extended Kalman filter uses the forecast state and its covariance $[x(t_F), P(t_F)]$ and the observations and their covariance $[y, R]$ to produce an optimal ("analyzed") estimation of the model state and its covariance $[x_A(t_F), P_A(t_F)]$:

$$x_A(t_F) = x(t_F) + P(t_F) H^T [R + H P(t_F) H^T]^{-1} \times [y - H x(t_F)]$$

and

$$P_A(t_F) = P(t_F) - P(t_F) H^T [R + H P(t_F) H^T]^{-1} H P(t_F).$$

(14)

The computational expense of the Kalman filter (14) is extremely large because one needs to invert the matrix $R + HPH^T$ and apply the tangent linear model to each column and the adjoint model to each row of the covariance matrix (Fisher 2001). The commonly used method to reduce the computational cost is to propagate (only) the projection of the covariance matrix onto a low-dimensional subspace (span $\{s_1, \ldots, s_n\}$). The ensemble Kalman filter (Houtekamer and Mitchell 2001) uses a Monte Carlo approach to define this subspace and to approximate the time-evolving covariance matrix.

The subspace (i.e., the ensemble of perturbations at the analysis time $t_F$) should contain the directions $s_k(t_F)$ along which the error has the maximal growth. Consequently the initial ensemble should be defined based on the singular vectors $s_k(t_0)$. This approach is used at the European Centre for Medium-Range Weather Forecasts (ECMWF) to generate initial perturbations for ensemble forecasts (Buizza et al. 2000; Buizza and Palmer 1995; Hamill et al. 2003).

c. Targeted observations

Adaptive observations placed in well-chosen locations can reduce the initial condition uncertainties and decrease forecast errors. A number of methods were proposed to "target observations," that is, to select areas where additional observations are expected to improve considerably the skill of a given forecast. Singular vectors identify sensitive regions of the atmospheric flow and can be used to optimally configure the observational network.
Singular vectors can identify the most sensitive regions of the atmosphere for targeted observations as long as the linearity assumption of error propagation holds (Hansen and Smith 2000). Majumdar et al. (2002) compare the singular vector approach for observation targeting with the ensemble transform Kalman filter. Palmer et al. (1998) argue that for predictability studies an appropriate metric is the perturbation energy. Daescu and Navon (2004) discuss the adaptive observation problem in the context of 4DVAR data assimilation. Leutbecher et al. (2002) and Leutbecher (2003) use Hessians to optimally place the adaptive observations.

3. Computation of singular vectors for chemical models

In this section we discuss the computational challenges associated with the chemical singular vectors and propose an approach for calculating them accurately. A numerical eigenvalue solver applied to (8) requires a symmetric matrix $M^*BM$ in order to successfully employ Lanczos iterations, and to guarantee that the numerical eigenvalues are real. There are two approaches to computing adjoints: continuous and discrete. In the continuous approach the adjoint of the continuous differential equations is derived, then solved numerically. In the discrete approach the numerical solution is (considered to be) the forward model and its adjoint is constructed. The symmetry requirement imposes the use of the discrete adjoint $M^*$ of the tangent linear operator $M$ in (8). The computation of discrete adjoints for stiff systems is a nontrivial task (Sandu et al. 2003). In addition, computational errors (which can destroy symmetry) have to be small.

For a given model a symmetry indicator is constructed based on two random perturbation vectors $u(t_0)$ and $v(t_0)$ that are propagated forward in time:

$$u(\tau) = M_{t_0\rightarrow \tau}u(t_0) \quad \text{and} \quad v(\tau) = M_{t_0\rightarrow \tau}v(t_0).$$

(15)

The symmetry residual is the difference:

$$r(\tau) = \langle u(\tau), M^*_{t_0\rightarrow \tau}M_{t_0\rightarrow \tau}v(\tau) \rangle - \langle v(\tau), M^*_{t_0\rightarrow \tau}M_{t_0\rightarrow \tau}u(\tau) \rangle.$$

(16)

If $M^*$ is the exact discrete adjoint of $M$ then $r(\tau) = 0$ for all $\tau$. However, both $M$ and $M^*$ are evaluated numerically and in practice we expect the symmetry residual $r(\tau)$ to have small (but nonzero) values.

To illustrate possible problems with losing the symmetry we consider the SAPRC-99 atmospheric gas-phase reaction mechanism (Carter 2000), which has 93 species and 235 reactions. The forward, tangent linear, and adjoint models are implemented using the Kinetic Preprocessor (KPP), an automatic code generator (Damian et al. 2002; Daescu et al. 2003; Sandu et al. 2003). Several numerical experiments revealed that the magnitude of the symmetry residual depends on the choice of numerical integrator. Among the Rosenbrock integrators available in KPP we selected Rodas4 (Sandu et al. 2003), which performs best with respect to symmetry. The variation of $r(\tau)$ with time is shown in

Fig. 1. (a) The variation of the symmetry residual with time shows a loss of symmetry during the initial transient of the stiff ODE. (b) The symmetry residual decreases rapidly with the number of projection iterations.
Fig. 1a (solid line). Surprisingly, the symmetry is lost during the stiff transient at the beginning of the integration interval, where the symmetry residual jumps from $10^{-16}$ to $10^{-2}$.

The loss of symmetry is due to the stiffness of the chemical terms. To understand this behavior we consider a singular perturbation model for the chemical system of the form

$$\frac{dy}{dt} = f(y, z), \quad \epsilon \frac{dz}{dt} = g(y, z), \quad \epsilon \ll 1. \quad (17)$$

A singular perturbation problem (17) is a model ordinary differential equation that explicitly separates the dynamics of the slow component $y$ and of the fast component $z$. This model problem is widely used in the theoretical study of the behavior of stiff systems and of the stiff numerical methods (Hairer and Wanner 2004).

Perturbations propagate through the tangent linear model of (17):

$$\frac{d\delta y}{dt} = f_s(y, z)\delta y + f_z(y, z)\delta z \quad \text{and}$$

$$\epsilon \frac{d\delta z}{dt} = g_s(y, z)\delta y + g_z(y, z)\delta z \quad (18)$$

For $\epsilon \to 0$, the perturbation vectors in (18) are of the form

$$\delta z = -g_z^{-1}(y, z)g_s(y, z)\delta y. \quad (19)$$

During the numerical computation of the eigenvectors, ARPACK (or any other solver) generates vectors $x = [\delta_{i0}, \delta_{i1}]^T$, which do not satisfy (19). These vectors are the initial conditions for the tangent linear model and are propagated forward, then backward through the adjoint model, in order to evaluate the matrix-vector products $M^*BMx$. Strong, artificial transients appear in the tangent linear model because of the fact that the initial perturbations are away from the slow manifold described by (19).

To correct this we apply the tangent linear model on the initial perturbation for a short time, which is equivalent to “projecting” the initial perturbation onto the slow evolution manifold in (19). The result is then used to initialize the subsequent tangent linear model run. To preserve operator symmetry, another projection using the adjoint model needs to be performed at the end of the adjoint integration. Consequently the matrix-vector products are computed as

$$w = \Pi^*M^*_pM_{y_0-x_i}BM_{y_0-x_i}^*\Pi x, \quad (20)$$

where $\Pi$ and $\Pi^*$ denote the projection operations performed with the tangent linear and the adjoint models respectively. Note that in the 3D model the projections will be performed at the beginning and at the end of each time split interval.

Numerical tests revealed that a small number of projection steps are sufficient in practice to substantially enhance symmetry. Figure 1b presents the evolution of the symmetry residual with the number of projection steps. The symmetry is markedly improved after only two projection steps.

Figure 1a (dashed lines) presents the evolution of the symmetry residual when six projection steps are performed with the very small step size of $10^{-9}$ s. The symmetry error during the stiff transient is reduced to $10^{-11}$. Note that the projection time step is of the order of the fastest scales in the system, and is much smaller than the numerical integration time step. We next extend these results to 3D chemical transport models.

4. 3D chemical transport models

Chemical transport models solve the mass-balance equations for concentrations of trace species in order to determine the fate of pollutants in the atmosphere. In this section we briefly describe the governing mass-balance equations and the tangent linear and adjoint models; a detailed discussion is given in Sandu et al. (2005). Tangent linear and adjoints of transport models are discussed in Daley (1995) and Vukicevic and Hess (2000).

Let $c_i$ be the mole-fraction concentration of chemical species $i$, $Q_i$ the rate of surface emissions, $E_i$ the rate of elevated emissions, $V_{i_{dep}}$ the deposition velocity, and $f_i$ the rate of chemical transformations. Furthermore, the inflow, outflow, and ground boundaries of the computational domain are denoted by $\Gamma_{in}$, $\Gamma_{out}$, and $\Gamma_{ground}$, respectively; $u$ is the wind field vector; $K$ is the turbulent diffusivity tensor; and $\rho$ is the air density. The evolution of $c_i$ is described by the following equations:

$$\frac{\partial c_i}{\partial t} = -u \cdot \nabla c_i + \frac{1}{\rho} \nabla \cdot (\rho K \nabla c_i) + \frac{1}{\rho} f_i(p) + E_i,$$

$$t_0 \leq t \leq t_F,$$

$$c_i(t_0, x) = c_{i0}^0(x),$$

$$c_i(t, x) = c_{i0}^{in}(t, x) \quad \text{for} \quad x \in \Gamma_{in},$$

$$K \frac{\partial c_i}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma_{out},$$

$$K \frac{\partial c_i}{\partial n} = V_{i_{dep}} c_i - Q_i \quad \text{for} \quad x \in \Gamma_{ground},$$

$$1 \leq i \leq N_{spec}. \quad (21)$$

We refer to (21) as the forward model.
A perturbation $\delta c^0$ of the initial conditions will result in perturbations $\delta c(t)$ of the concentration field at later times. The evolution of these perturbations is governed by the equations:

$$\frac{\partial \delta c_i}{\partial t} = -\mathbf{u} \cdot \nabla \delta c_i + \frac{1}{\rho} \nabla \cdot (\rho \mathbf{K} \nabla \delta c_i) + F_i \rho(\mathbf{c}) \delta c,$$

$$t_0 \leq t \leq t_F$$

$$\delta c_i(t_0, x) = \delta c_i^0(x),$$

$$\delta c_i(t, x) = 0 \quad \text{for} \quad x \in \Gamma^{in},$$

$$K \frac{\partial \delta c_i}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^{out}, \quad \text{and}$$

$$K \frac{\partial \delta c_i}{\partial n} = \mathbf{V}_{i \text{dep}} \delta c_i \quad \text{for} \quad x \in \Gamma^{ground},$$

for all $1 \leq i \leq N_{\text{spec}}$. (22)

Equation (22) is referred to as the tangent linear model associated with the forward model (21). Here $\mathbf{F} = \partial f / \partial \mathbf{c}$ denotes the Jacobian of the chemical rate function $f$, and $F_i$ is its $i$th row.

The continuous adjoint model associated with the forward model (21) [or, more exactly, the adjoint of the tangent linear model (22)] describes the evolution of the adjoint variables $\lambda$:

$$\frac{\partial \lambda_i}{\partial t} = -\nabla \cdot (\mathbf{u} \lambda_i) - \nabla \cdot \left( \frac{\rho \mathbf{K} \lambda_i}{\rho} \right)$$

$$- [\mathbf{F} \rho(\mathbf{c}) \lambda_i] \rightarrow \phi_i, \quad t_F \geq t \geq t_0$$

$$\lambda_i(t_F, x) = \lambda_i^f(x),$$

$$\lambda_i(t, x) = 0 \quad \text{for} \quad x \in \Gamma^{in},$$

$$\lambda_i + \rho \mathbf{K} \frac{\partial (\lambda_i / \rho)}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^{out}, \quad \text{and}$$

$$\rho \mathbf{K} \frac{\partial (\lambda_i / \rho)}{\partial n} = \mathbf{V}_{i \text{dep}} \lambda_i \quad \text{for} \quad x \in \Gamma^{ground},$$

for all $1 \leq i \leq N_{\text{spec}}$. (23)

The ground boundary condition is obtained from the fact that $\mathbf{u} \cdot \mathbf{n} = 0$ at ground level. The forcing function $\phi_i$ depends on the particular cost functional under consideration (Sandu et al. 2005).

The numerical solution operator for (21) is based on an operator splitting approach, where the transport and the chemistry steps are taken successively. With $T$ the numerical solution operator for transport, and $C$ the solution operator for chemistry, one step of the solution operator reads

$$\mathbf{c}^{n+1} = T \mathbf{c}^n C \mathbf{c}^n T \mathbf{c}^n \mathbf{c}^n.$$

The tangent linear model of (24) is constructed from the tangent linear transport $T$ and chemistry $C$ operators. As explained in section 3 a projection onto the chemical slow manifold $\Pi$ is applied before each linearized chemistry:

$$\delta \mathbf{c}^{n+1} = T \mathbf{c}^n C \mathbf{c}^n T \mathbf{c}^n \mathbf{c}^n.$$

The discrete adjoint model is based on the discrete adjoints of the transport $T^*$ and chemistry $C^*$ numerical schemes. A chemical adjoint projection $\Pi^*$ is applied after each adjoint chemistry step:

$$\lambda^m = \mathbf{C}^* \mathbf{T}^* \lambda^m + \phi^m. \quad (26)$$

The projection $\Pi^*(\Pi^*)$ is applied at every operator split step, since the transport step perturbs the chemical quasi equilibria.

5. Error norms

In numerical weather prediction models, variables have different physical units (wind velocity, temperature, air density, etc). The energy norms provide a unified measure for the magnitude of perturbations in variables of different dimensions.

In chemical transport models, variables are concentrations of chemical species. Since all variables have the same physical meaning, and similar units, we expect that simple $L^2$ norms will provide a reasonable measure of the magnitude of the perturbation

$$\| \delta \mathbf{c} \|^2 = \sum_{i,j,k} \sum_{s=1}^{N_{\text{spec}}} (\delta c_{i,j,k}^s)^2 = \langle \delta \mathbf{c}, \mathbf{A} \delta \mathbf{c} \rangle \quad \text{with} \quad \mathbf{A} = \mathbf{I}.$$

Here $c_{i,j,k}^s$ denotes the concentration of chemical species $s$ at the grid point $(i, j, k)$ in the discrete model, and $\delta c_{i,j,k}^s$ is its perturbation.

In practice we are interested to assess the influence of perturbations onto a well-defined optimization area, and a given set of chemical species. In this case the perturbation magnitude at the final time is defined in terms of a local projection operator (Buizza 1994), that is, a matrix with the diagonal entries equal to one for the selected optimization variables and equal to zero elsewhere:

$$\| \delta \mathbf{c}(t_F) \|^2_B = \sum_{\text{area}} \sum_{\text{spec}} (\delta c_{i,j,k}^s)^2$$

$$= \langle \delta \mathbf{c}(t_F), \mathbf{B} \delta \mathbf{c}(t_F) \rangle \quad \text{with} \quad \mathbf{B} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$
Because concentrations of different species vary by many orders of magnitude we expect that the perturbations of the more abundant species (e.g., CO) will dominate the total error norm in (27). To have a balanced account for the influence of all species it is of interest to consider the directions of maximal relative error growth (i.e., the directions that maximize):

\[
\sigma^2 = \frac{\delta c(t_f)}{c(t_f)} B \frac{\delta c(t_f)}{c(t_f)}
\]

This can be accomplished by using the logarithms of the concentrations \(\log c_{i,j,k}^s\) as model variables. In practice it is advantageous to approximate the relative errors by the absolute errors \(\delta c_{i,j,k}^s\) scaled by “typical” concentration values \(w_{i,j,k}^s\).

\[
\frac{\delta c_{i,j,k}^s(t_f)}{c_{i,j,k}^s(t_f)} \approx \frac{\delta c_{i,j,k}^s(t_f)}{c_{i,j,k}^s(t_f)} \approx W_{t_f}^{-1} \delta c(t_f),
\]

where

\[
W_t = \text{diag}[w_{i,j,k}^s(t)],
\]

and maximize

\[
\sigma^2 = \frac{\langle W_{t_f}^{-1} \delta c(t_f), B W_{t_f}^{-1} \delta c(t_f) \rangle}{\langle W_{t_0}^{-1} \delta c(t_0), A W_{t_0}^{-1} \delta c(t_0) \rangle}.
\]

**Fig. 2.** The dominant eigenvalues \((\lambda = \sigma_i^2)\) for the 12-, 24-, and 48-h simulations. The rapid decrease in magnitude indicates that uncertainty in the optimization region can be captured by only a few singular vectors.

**Fig. 3.** The 2-km level wind fields during the 12-, 24-, and 48-h simulations starting at 0000 UTC 1 Mar 2001. The optimization area is shaded.
One reason for this approximation is that the typical concentrations $w_{ij,k}$ can be chosen to be bounded away from zero. More importantly, having the weights independent of the system state $c$ keeps the maximization problem in (29) equivalent to a generalized eigenvalue problem 

$$M^* t F \rightarrow t F W t F / H 11002$$

$$M t 0 \rightarrow t F W t 0 / H 11002$$

For the important case where $A = I$ the generalized eigenvalue problem in (30) is equivalent to the following eigenvalue problem:

$$W_0 M^* W_0 W_0 B W_0 M_0 t x_0 W_0 v = \sigma^2 v$$

where

$$v = W_0^{-1} s_k(t_0).$$

Fig. 4. Dominant singular vectors (SV: O₃ sections) for the 12-, 24-, and 48-h simulations starting at 0000 UTC 1 Mar 2001. The optimized criterion is the ground-level O₃ value in the shaded area.
The weight matrices $W$ can be different at the initial and at the final time in order to account for different "typical concentration" levels at nighttime versus daytime.

6. Numerical results

The numerical tests use the state-of-the-art regional atmospheric chemical transport model (STEM; Carmichael et al. 2003). The simulation covers a region of $7200 \text{ km} \times 4800 \text{ km}$ in East Asia and the simulated conditions correspond to March 2001. More details about the forward model simulation conditions and comparison with observations are available in Carmichael et al. (2003).

The computational grid has $N_x \times N_y \times N_z$ nodes with $N_x = 30$, $N_y = 20$, $N_z = 18$, and a horizontal resolution of $240 \text{ km} \times 240 \text{ km}$. The chemical mechanism is SAPRC-99 (Carter 2000), which considers the gas-phase atmospheric reactions of volatile organic and nitrogen oxides in urban and regional settings. The meteorological fields have been computed using the Regional Atmospheric Modeling System (more information available online at http://rams.atmos.colostate.edu/), and analyzed offline (data assimilation of the meteorological observations has been performed before the chemical transport simulations). The initial and boundary conditions have been obtained from a long run of the model before the start time of the current computations; details can be found Carmichael et al. (2003). While the simulations described in Carmichael et al. (2003) use a grid resolution of $80 \text{ km} \times 80 \text{ km}$, in the current paper we use a coarser grid in order to reduce the CPU time needed by the singular vector calculations.

The adjoint of the comprehensive model STEM is discussed in detail in Sandu et al. (2005). Both the forward and adjoint chemical models are implemented using KPP (Damian et al. 2002; Daescu et al. 2003; Sandu et al. 2003). The forward and adjoint models are parallelized using PAQMSG (Miehe et al. 2002). PARPACK (available online at http://www.caam.rice.edu/~kristyn/parpack_home.html) was used to solve the symmetric generalized eigenvalue problems.

The singular vectors $s(N_x, N_y, N_z, N_{\text{spec}})$ in (8) are represented by four-dimensional arrays. To visualize them we separately consider the vector sections corresponding to different chemical species. Further, each three-dimensional section is reduced to a two-dimensional "top" view by adding the values in each vertical column, or to a two-dimensional "south" view by adding the values in each north–south column.

Numerical results for different optimization regions,
optimization species, simulation intervals, meteorological data, and error norms are presented next.

a. Singular vectors for different simulation intervals

We first consider the case where the optimization criterion is the ground-level ozone concentration in a 720 km × 960 km area covering Korea. The singular vector analysis presented next will help answer the following questions:

- In which areas will small changes in the initial conditions grow fastest to impact the ozone levels over

**Fig. 6.** Same as in Fig. 4, but for third and fourth dominant SVs (O$_3$ sections).
Korea after 12, 24, and 48 h and what is their rate of growth?

- How should the initial perturbations be constructed for ensemble simulations in order to properly describe the uncertainty in ground ozone predictions over Korea after 12, 24, and 48 h?
- Where are additional observations needed the most in order to improve 12-, 24-, and 48-h predictions of ground-level Korean ozone?

The largest 12 singular values for 12-, 24-, and 48-h simulations started at 0000 UTC 1 March 2001 are...
shown in Fig. 2. The rapid decrease of the eigenvalue magnitude indicates that the uncertainty in the ground-level ozone within the optimization region can be captured with only a few singular vectors. The eigenvalues decrease faster for longer simulation intervals.

The meteorological conditions are an important factor in determining the singular vectors. The 2-km level wind fields for the simulation interval under consideration are shown in Fig. 3.

The top and south views for O₃ sections of the dominant et al. for the 12-, 24-, and 48-h simulations are presented in Fig. 4. Singular vectors are localized near
the optimization area in both the horizontal and the vertical directions. As expected, for longer simulation intervals the singular vectors spread further away from the optimization region. The singular vectors are not confined to the lowest layers, but also show important regions located between 1 and 3 km. This is because of the transport processes, which exchange material from the surface into the free troposphere, and which bring free tropospheric air back to the surface. This has important implications for the design of measurement sys-

**Fig. 9.** The evolved dominant SVs after the 24-h simulation (i.e., at 0000 UTC 2 Mar 2001).
tems, since it shows that surface measurements alone are not sufficient for a correct representation of ground-level concentrations.

Several dominant singular vectors for the 12-, 24-, and 48-h simulations starting at 0000 UTC 1 March 2001 are shown in Figs. 4, 5, and 6 (O₃ sections), Fig. 7 (NO₂ sections), and Fig. 8 (HCHO sections). The NO₂ and HCHO are important species involved in the photochemical production of ozone. They also are species that are directly emitted into the atmosphere as a result of combustion processes.

A close look at the structure of the dominant singular vectors reveals the following:

• Since different singular vectors are orthogonal they contain different information about the areas of maximal error growth.
• The eigenvectors evolve in time as the length of the simulation interval increases. They tend to expand farther away from the optimization area, illustrating that perturbations in a wider area at earlier times impact the optimization area.
• The shapes and the magnitudes of the O₃, NO₂, and HCHO sections show subtle differences, illustrating the different influences that these species have on ground-level O₃ after 12, 24, and 48 h.

Singular vectors contain valuable information for the placement of additional observations in order to improve predictions of ground-level Korean ozone. According to this analysis, additional O₃ measurements may have to be placed in a different location than additional NO₂ or HCHO observations. Moreover, the optimal location of observations changes in time and drifts away from the optimization area for longer intervals. In conclusion, what is needed is a well thought out distribution of sites measuring many parameters simultaneously.

b. Evolved singular vectors

The perturbations initialized along each dominant singular vector develop in time becoming “evolved singular vectors.” We are interested in the shape of these perturbations at the end of the 24-h simulation interval.

The evolved singular vectors (scaled to have the A-norm equal to 1) are displayed in Fig. 9. The largest values of singular vectors are clustered above the optimization region. This is expected since the SVs are constructed to optimize final-time perturbation norm in the optimization region.

c. The linearity assumption

Inherent in the singular vector calculation is the assumption that small perturbations propagate according to the tangent linear model dynamics (Tanguay et al. 1997). To assess the validity of this linearity assumption we perturb the initial state with scaled versions each of the first five singular vectors. The scaling is chosen such that the ground-level ozone perturbations are $\sim 5\%$–10% of the reference ozone values. The perturbed initial state is propagated forward for 24 h using the full, nonlinear model. The perturbation at the final time is the difference between the perturbed and the reference final states. The B norms of the evolved perturbations are divided by the A norms of the initial perturbations. The results shown in Table 1 reveal that the perturbation magnitude ratios approximate well the singular values.

The structure of the nonlinearly evolved perturbations at final time are shown in Fig. 10. The evolved perturbation structure is similar to that of the evolved singular vectors shown in Fig. 9. Since both the magnitude and the structure of the linearly evolved perturbations match those of the nonlinearly evolved perturbations we conclude that the linearity assumption holds (at least) for the 24-h simulation interval under consideration.

We next consider a random perturbation vector $\mathbf{r}$ with components drawn from a uniform distribution with amplitude $\pm 10\%$ of the initial concentrations $r_{i,k}(t_0) \in \mathbb{U}[0.9r_{i,k}(t_0), 1.1r_{i,k}(t_0)]$. The perturbation evolved for 24 h and its B norm was taken. The perturbation components along each of the singular vectors $\mathbf{r}_i(t_0) = \sigma_i r(t_0), \mathbf{A}_s_i(t_0) \mathbf{s}_i(t_0)$ evolve into

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
Initial perturbation & Singular values & Ratio of perturbation energies \\
\hline
Along $s_i(t_0)$ & $\sigma_i^2 = 9.494 \times 10^{-2}$ & $9.685 \times 10^{-2}$ \\
Along $s_i(t_0)$ & $\sigma_i^2 = 5.146 \times 10^{-2}$ & $5.169 \times 10^{-2}$ \\
Along $s_i(t_0)$ & $\sigma_i^2 = 3.982 \times 10^{-2}$ & $3.944 \times 10^{-2}$ \\
Along $s_i(t_0)$ & $\sigma_i^2 = 2.499 \times 10^{-2}$ & $2.573 \times 10^{-2}$ \\
Along $s_i(t_0)$ & $\sigma_i^2 = 1.756 \times 10^{-2}$ & $1.741 \times 10^{-2}$ \\
Random $\mathbf{r}$ & $\Sigma_{i=1}^{12} \sigma_i^2 r(t_0), \mathbf{A}_s(t_0)^2 = 3.174 \times 10^{-22}$ & $3.210 \times 10^{-22}$ \\
\hline
\end{tabular}
\end{center}
The total B norm of the first \( n \) components of the perturbation is \( \sum_{i=1}^{n} \sigma_{i}(t_{0}) \). The results in Table 1 (last row) show that the evolved perturbation components along the first 12 singular vectors account for virtually all the B norm of the random perturbation at the final time. This result confirms the fact that perturbation effects can be captured using only a small subspace of dominant singular vectors.

d. Singular vectors versus adjoints

To illustrate the difference between the information conveyed by the singular vectors and by the adjoint.
variables we consider again the ground-level O\textsubscript{3} in the Korea optimization area and focus on the 24-h simulation starting at 0000 UTC 1 March 2001. The cost function in the adjoint calculation is the sum of squared ground-level O\textsubscript{3} concentrations in the optimization area. The adjoint variables are computed through a 24-h backward integration and are shown in Fig. 11.

To assess the relationship between the adjoint variable $\lambda(t_0)$ and different singular vectors $s_k(t_0)$ we consider the {correlation coefficients} $\rho_k = \langle \lambda(t_0), A s_k(t_0) \rangle / (\|\lambda(t_0)\|_A \|s_k(t_0)\|_A)$. Specifically, we compute the correlations between individual (and homologous) sections of $\lambda(t_0)$ and $s_k(t_0)$. The results are shown in Fig. 12. For all sections the correlation of the adjoint and the first singular vector is the strongest. The O\textsubscript{3} section of the adjoint in particular is very weakly correlated with the remaining singular vectors.

The comparison of adjoint variables with the singular vectors (Figs. 4–8) points to the following conclusions.

- The adjoints have a similar structure with the first singular vectors. The next dominant singular vectors (the second, the third, etc.) carry additional information about the areas where changes have a high impact on the optimization area. This additional information is not captured by the adjoint.

![Adjoints for the 24-h simulation starting at 0000 UTC 1 Mar 2001.](image1)

**Fig. 11.** Adjoints for the 24-h simulation starting at 0000 UTC 1 Mar 2001.

![Correlations between homologous sections of the adjoint variable and of the dominant SVs.](image2)

**Fig. 12.** Correlations between homologous sections of the adjoint variable and of the dominant SVs.
The adjoint covers a wider area following the flow pattern, while the singular vectors remain localized, even for larger simulation interval.

These findings are in good agreement with the results reported for meteorological models by Gelaro et al. (1998).

e. Influence of the meteorological conditions

To assess the influence of different meteorological conditions on the singular vectors we perform a 24-h simulation starting at 0000 UTC 26 March 2001, for the same optimization criterion (ground-level O\(_3\) in the Korea area). The 2-km level wind fields on 26 March are shown in Fig. 13. A comparison with the conditions present on 1 March (shown in Fig. 3) reveals that the meteorological conditions were considerably different during these two days.

The O\(_3\), NO\(_2\), and HCHO sections of the four dominant singular vectors are shown in Fig. 14. There are clear differences between the structure of the singular vectors at 26 and 1 March (Figs. 4–8). This shows the important role that meteorology plays in the distribution of pollutant concentrations.

f. Influence of the optimization region

In the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain. We now analyze how the choice of the optimization region impacts the singular vectors.

The O\(_3\), NO\(_2\), and HCHO sections of the dominant eigenvector for another 24-h, 1 March simulation are shown in Fig. 15. The optimization criterion is ground-level O\(_3\) over a region of the same area, but located in southeast China (the gray area on the map). As expected, singular vectors are localized over the optimization area.

Another numerical test is performed for a optimization area that covers 24 grid cells over Japan, Korea, and southeast China. The magnitude of the largest eigenvalues decreases at a slower rate (as shown in Fig. 18). This is because of the optimization area being larger than in the previous numerical experiments. About 30 eigenvalues are needed for a two orders of magnitude decrease in the magnitude of the eigenvalues; therefore about 30 singular vectors are needed to accurately capture the uncertainty. The O\(_3\), NO\(_2\), and HCHO sections of the dominant singular vectors are shown in Fig. 16. Singular vectors are localized over China and Korea; there are no lobes localized over Japan. Because of the westerly flows changes in initial concentration fields over Japan do not impact significantly ground O\(_3\) concentration after 24 h over the optimization region.

To further show the influence of the optimization region we consider a very large area with over 100 cells covering parts of China, Korea, and Japan (the gray area on the map in Fig. 17). The magnitude of the largest 40 eigenvalues is shown in Fig. 18. The decrease of eigenvalue magnitude is slower for the larger regions, and therefore, more eigenvectors are needed to capture the uncertainty. The ratio of the smallest to largest computed eigenvalues is \(\lambda_1/\lambda_{40} = 22\) for the 100-cell region, compared with \(\lambda_1/\lambda_{40} = 472\) for the 24-cell region. The O\(_3\), NO\(_2\), and HCHO sections of the dominant singular vectors are shown in Fig. 17. The dominant singular vectors are also localized to Korea and southeast China. Note that the first few dominant singular vectors are insufficient to accurately describe the uncertainty.
The singular vectors also depend on the choice of error norms at the initial \(A\) and final time \(B\). Additional numerical tests (not shown here) have been performed using the ground-level concentrations of 66 long-lived species over Korea. The new singular vectors displayed clear differences from the ones computed based on ozone ground level over Korea.

7. Conclusions

In this work we study the computational aspects of singular vector analysis of chemical transport models. Singular vectors span the directions of maximal error growth in a finite time, as measured by specific error norms.
To maintain the symmetry of the tangent linear–adjoint operator $M^*M$ it is necessary to employ discrete adjoints. A projection method is proposed to preserve the symmetry of $M^*M$ operators for stiff chemical systems. The application of this technique is extended to 3D chemical transport models. Different definitions of the perturbation error norms are discussed.

Numerical results are presented for a 3D chemical transport simulation of atmospheric pollution in East Asia in March 2001. The assumption of linear propagation of perturbations, intrinsic in the singular vector calculation, was checked numerically for a 24-h simulation interval. The singular values and the structure of the singular vectors depend on the length of the simu-

**Fig. 15.** Dominant SVs for the 24-h simulation starting at 0000 UTC 1 Mar 2001. The optimized criterion is ground-level $O_3$ over the gray region in southeast China.
lation interval, the meteorological data, the location of optimization region and the selection of optimized species, the choice of error norms, and the size of the optimization region.

While in the study of atmospheric dynamics the dominant singular vectors are associated with unstable modes, in the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain (e.g., where the model prediction needs to be improved). In this paper we illustrate the use of singular vectors to describe uncertainties in the initial conditions. Other very impor-
tant sources of uncertainty in air quality models that need to be quantified in real applications are the emissions, the meteorological fields, the deposition velocities, and the top and lateral boundary conditions for regional models.

The predictions of air quality models are corrupted by uncertainties coming from the initial distribution of the chemical fields, the boundary conditions, the rates of emission of pollutants, and the meteorological fields. In this paper we illustrate the use of singular vectors to quantify the propagation of uncertainties from the initial conditions; but the other sources need to be accounted for in a real data assimilation setting. The decrease of the singular values for longer simulation in-

Fig. 17. Dominant SVs for the 24-h simulation over a large optimization area.
Fig. 18. The dominant eigenvalues for simulations on 24-cell and large optimization areas.

Intervals is a result of the fact that, as time progresses, the final solution is driven more by emissions and less by the initial conditions. Consequently, the effect of uncertainties in emission sources on the final state becomes more important.

Most of the uncertainty in the optimization region at the final time is determined by the uncertainty along the dominant singular vectors at the initial time. The uncertainty (error) growth rates along each direction are given by the corresponding singular values. For limited optimization regions the singular values decrease rapidly, and a few dominant singular vectors are sufficient to capture most of the uncertainty. For large optimization regions the singular values decrease slowly. The areas of influence are no longer localized, and uncertainty from all over the computational domain contributes to the uncertainty in the optimization area at the final time. As a consequence for data assimilation, small ensembles are sufficient if the observations are localized, or if one seeks improved predictions over a relatively small, well-defined region.

To improve predictions within the optimization region additional observations are needed in the areas described by the dominant singular vectors. Additional O$_3$ measurements have to be placed in a different location than additional NO$_2$ or HCHO observations. The optimal location of observations changes in time and drifts away from the optimization area for longer intervals.

The dominant singular vector has a similar structure to the adjoint variable. The next singular vectors carry additional information about the high sensitivity areas, which is not captured by a simple adjoint analysis.

The computation of singular vectors is computationally intensive. In our experiments 40–100 iterations were necessary for PARPACK to converge (taking between 8–16 h of CPU time for a parallel run on 30 Opteron processors). Each iteration includes one forward and tangent linear model run and one adjoint run. The cost of one forward and tangent linear model (using a direct-decoupled approach and reusing the matrix factorizations) followed by one backward adjoint integration is less than 3 times the cost of the forward trajectory calculation (Sandu et al. 2005). The calculation of singular vectors is at least as expensive as a full 4DVAR data assimilation cycle, where 20–30 iterations are typically sufficient for a substantial decrease in the cost function (Sandu et al. 2005). Note that the cost of performing the chemical projections accounts for only a small percent of the total computational time.

As the field of chemical weather forecasting grows, it can be anticipated that singular vectors will find many applications. The results presented in this paper are a first step in this direction.

Acknowledgments. This work was supported by the National Science Foundation through Award NSF ITR AP&IM 0205198. Sandu’s work was also partially supported by the Award NSF CAREER ACI-0413872. We thank Virginia Tech’s Laboratory for Advanced Scientific Computing (LASC) for the use of the Anantham cluster. The authors also thank the anonymous reviewers for their constructive comments, which helped improve this work.

REFERENCES


Carter, W., 2000: Implementation of the SAPRC-99 chemical mecha-


