Lecture 2:
The Mechanics of 4D-Var
Outline

• The conjugate gradient algorithm
• Preconditioning
• Covariance modeling
• Background quality control
The Conjugate Gradient Algorithm
(cgradient.h & congrad.F)

Recall the incremental cost function:

\[ J = \frac{1}{2} \delta z^T D^{-1} \delta z + \frac{1}{2} (G \delta z - d)^T R^{-1} (G \delta z - d) \]

\[ = \frac{1}{2} \delta z^T \left( D^{-1} + G^T R^{-1} G \right) \delta z - \delta z^T G^T R^{-1} d + \frac{1}{2} d^T R^{-1} d \]

At the minimum of \( J \) we have \( \partial J / \partial \delta z = 0 \)

\( \left( D^{-1} + G^T R^{-1} G \right) \delta z - G^T R^{-1} d = 0 \)

i.e. solve \( A \delta z = b \)
The Conjugate Gradient Algorithm

The ECMWF “congrad” of Fisher (1997) for inner-loop \( k+1 \):

\[
\delta \hat{Z}_k = \delta z_k + \tau_k h_k \quad \text{trial step}
\]

\[
\hat{g}_k = \frac{\partial J}{\partial \delta \hat{Z}_k} \quad \text{gradient @ trial step}
\]

\[
\alpha_k = -\tau_k h_k^T g_k / \left( h_k^T (\hat{g}_k - g_k) \right) \quad \text{optimum step}
\]

\[
\delta z_{k+1} = \delta z_k + \alpha_k h_k \quad \text{new starting point}
\]

\[
g_{k+1} = g_k + (\alpha_k / \tau_k) (\hat{g}_k - g_k) \quad \text{gradient @ new point}
\]

\[
\beta_{k+1} = g_{k+1}^T g_{k+1} / g_k^T g_k
\]

\[
h_{k+1} = -g_{k+1} + \beta_{k+1} h_k \quad \text{new descent direction}
\]
The Lanczos Connection

Cornelius Lanczos (1893-1974)
The Lanczos Connection

The CG algorithm is equivalent to:

\[ Aq_{k+1} = \gamma_{k+1}q_{k+2} + \delta_{k+1}q_{k+1} + \gamma_k q_k \]

"Lanczos recursion relation"

\[ q_k = g_k / \|g_k\|; \quad \delta_{k+1} = (1/\alpha_{k+1} + \beta_{k+1}/\alpha_k); \quad \gamma_k = -\beta_{k+1}^{1/2}/\alpha_k \]

\[ V_k = \left(\begin{array}{c} q_1 \\ \vdots \\ q_k \end{array}\right) \]

Orthonormal Lanczos vectors

\[ q_i^T q_j = \delta_{ij} \]

\[ V_k^T V_k = I_k \]

\[ AV_k = V_k T_k + \gamma_k q_{k+1} e_k^T \]

\[ T_k = \begin{pmatrix} \delta_1 & \gamma_1 \\ \gamma_1 & \delta_2 & \gamma_2 \\ \vdots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \gamma_{k-2} & \delta_{k-1} & \gamma_{k-1} \\ \gamma_{k-1} & \delta_k \end{pmatrix} \]
The Lanczos Connection

Specifically:

\[ A \approx V_k T_k V_k^T \]

\[ A^{-1} \approx V_k T_k^{-1} V_k^T \]
The Lanczos Connection

Gain (primal form):

\[ K = (D^{-1} + G^T R^{-1} G)^{-1} G^T R^{-1} \]

Practical gain matrix:

\[ \tilde{K}_{kk} = V_k T_k^{-1} V_k^T G^T R^{-1} \]

Useful for diagnostic applications (Lecture 5)
(The Lanczos vectors are in ADJname)
Preconditioning seeks to cluster the eigenvalues of $A$ via a transformation of variable $\mu$.
Preconditioning

At the minimum of $J$ we have $\partial J/\partial \delta z = 0$

\[
\left( D^{-1} + G^T R^{-1} G \right) \delta z - G^T R^{-1} d = 0
\]

i.e. solve $A \delta z = b$

Minimize:

\[
J = \frac{1}{2} \delta z^T A \delta z - \delta z^T b + c
\]

Introduce a new variable: $v = A^{1/2} \delta z$

\[
J = \frac{1}{2} v^T v - v^T A^{-T/2} b + c
\]

At the minimum: $\partial J/\partial v = v - A^{-T/2} b = 0$
Preconditioning

Recall the incremental cost function:

\[ J = \frac{1}{2} \delta z^T D^{-1} \delta z + \frac{1}{2} (G \delta z - d)^T R^{-1} (G \delta z - d) \]

Introduce a new variable: \( v = D^{-1/2} \delta z \)

\[ J(v) = \frac{1}{2} v^T v + \frac{1}{2} (G D^{1/2} v - d)^T R^{-1} (G D^{1/2} v - d) \]

\[ = \frac{1}{2} v^T \left( I + D^{1/2} G^T R^{-1} G D^{1/2} \right) v - v^T D^{1/2} G^T R^{-1} d + \frac{1}{2} d^T R^{-1} d \]

At the minimum of \( J \) we have \( \partial J / \partial v = 0 \)

\[ \left( I + D^{1/2} G^T R^{-1} G D^{1/2} \right) v - D^{1/2} G^T R^{-1} d = 0 \]

i.e. solve \( \tilde{A} v = \tilde{b} \) then \( \delta z = D^{1/2} v \)
Preconditioning

Solve \[ \tilde{A}v = \tilde{b} \]

\[ \tilde{A} = \left( I + D^{1/2} G^T R^{-1} G D^{1/2} \right) \]

Has eigenvalues clustered around 1

\[ J(\delta z) \]  
\[ J(v) \]
The Conjugate Gradient Algorithm

cgradient.h in v-space to minimize $J(v)$

\[
\hat{v}_k = v_k + \tau_k h_k
\]

trial step

\[
\hat{g}_k = D^{T/2} \frac{\partial J}{\partial \hat{z}_k}
\]

gradient @ trial step

\[
\alpha_k = -\tau_k h_k^T g_k \left( h_k^T (\hat{g}_k - g_k) \right)
\]

optimum step

\[
v_{k+1} = v_k + \alpha_k h_k
\]

new starting point

\[
g_{k+1} = g_k + \left( \alpha_k / \tau_k \right) (\hat{g}_k - g_k)
\]

gradient @new point

\[
\beta_{k+1} = g_{k+1}^T g_{k+1} / g_k^T g_k
\]

new descent direction

\[
h_{k+1} = -g_{k+1} + \beta_{k+1} h_k
\]

project into state-space

\[
\delta z_{k+1} = D^{1/2} v_{k+1}
\]
The Lanczos Connection

Gain (primal form):

\[ K = D^{1/2} \left( I + D^{T/2} G^T R^{-1} G D^{1/2} \right)^{-1} D^{T/2} G^T R^{-1} \]

Practical gain matrix:

\[ \tilde{K}_k = D^{1/2} V_k^{-1} V_k^T D^{T/2} G^T R^{-1} \]

Useful for diagnostic applications (Lecture 5)
(The Lanczos vectors are in ADJName)
Covariance Modeling

Recall the incremental cost function:

\[ J = \frac{1}{2} \delta z^T D^{-1} \delta z + \frac{1}{2} (G \delta z - d)^T R^{-1} (G \delta z - d) \]

At the minimum of \( J \) we have \( \partial J / \partial \delta z = 0 \)

\[ \partial J / \partial \delta z = D^{-1} \delta z + G^T R^{-1} (G \delta z - d) \]

where \( D = \text{diag}(B_x, B_b, B_f, Q) \)
Covariance Modeling

\( B_x = \) initial condition \textit{prior} (or background) error covariance matrix
\( B_f = \) surface forcing \textit{prior} error covariance matrix
\( B_b = \) open boundary condition \textit{prior} error covariance matrix
\( Q = \) \textit{prior} model error covariance matrix

Each covariance matrix is factorized according to:

\[
B = K_b \Sigma C \Sigma^T K_b^T
\]

\( C = \) univariate correlation matrix
\( \Sigma = \) diagonal matrix of error standard deviations
\( K_b = \) multivariate balance operator (for \( B_x \) and \( Q \) only)
Correlation Models

$C$ is further factorized as:

$$C = \Lambda L_v^{1/2} L_h^{1/2} W^{-1} L_h^{T/2} L_v^{T/2} \Lambda^T$$

$W$ = diagonal matrix of grid box volumes
$L_h$ = horizontal correlation function model
$L_v$ = vertical correlation function model
$\Lambda$ = matrix of normalization coefficients

$L_h$ and $L_v$ are based on solutions of 2D and 1D pseudo diffusion equations respectively:

$$\frac{\partial \eta}{\partial t} - \kappa_h \nabla^2 \eta = 0 \quad \frac{\partial \eta}{\partial t} - \kappa_v \frac{\partial^2 \eta}{\partial z^2} = 0$$
Correlation Models

$C$ is further factorized as:

$$C = \Lambda L_v^{1/2} L_h^{1/2} W^{-1} L_h^{T/2} L_v^{T/2} \Lambda^T$$

$W$ = diagonal matrix of grid box volumes  
$L_h$ = horizontal correlation function model  
$L_v$ = vertical correlation function model  
$\Lambda$ = matrix of normalization coefficients

$L_h$ and $L_v$ are based on solutions of 2D and 1D pseudo diffusion equations respectively:

$$\frac{\partial \eta}{\partial t} - \kappa_h \nabla^2 \eta = 0 \quad \frac{\partial \eta}{\partial t} - \kappa_v \frac{\partial^2 \eta}{\partial z^2} = 0$$
Correlation Models

Correlation length, $L$:

$$L^2 \approx 2kT$$
Covariance Modeling

\[ C = \Lambda L_v^{1/2} L_h^{1/2} W^{-1} L_h^{T/2} L_v^{T/2} \Lambda^T \]

\( \Lambda \) ensures that the range of \( C \) is \( \pm 1 \).

Suppose that \( x \) is divided into a balanced and unbalanced contribution: \( x = x_b + x_u \)

Examples of balance relations: geostrophy, hydrostatic

\[ \left( B_x \right)_u = \Sigma C \Sigma^T \]

\[ B_x = K_b \left( B_x \right)_u K_b^T \]
The Balance Operator
(defined BALANCE_OPERATOR)

Following Weaver et al (2005):

The total state vector increments
\[ \delta x = \begin{bmatrix} \delta T \\ \delta S \\ \delta \zeta \\ \delta u \\ \delta v \end{bmatrix} \]

The unbalanced state vector increments (except for \( \delta T \))
\[ \delta \hat{x} = \begin{bmatrix} \delta T \\ \delta S_u \\ \delta \zeta_u \\ \delta u_u \\ \delta v_u \end{bmatrix} \]

\[ (B_x)_u = \langle \delta \hat{x} \delta \hat{x}^T \rangle \]
\[ \delta x = K_b \delta \hat{x} \]
\[ B_x = \langle \delta x \delta x^T \rangle = K_b \langle \delta \hat{x} \delta \hat{x}^T \rangle K_b^T \]
\[ = K_b (B_u)_x K_b^T \]
The Balance Operator

\[
\begin{align*}
\delta S &= K_{ST} \delta T + \delta S_u \\
\delta \zeta &= K_{\zeta p} \delta \rho + \delta \zeta_u \\
\delta u &= K_{up} \delta p + \delta u_u \\
\delta v &= K_{vp} \delta p + \delta v_u \\
\delta \rho &= K_{\rho T} \delta T + K_{\rho S} \delta S \\
\delta p &= K_{p\rho} \delta \rho + K_{p\zeta} \delta \zeta
\end{align*}
\]

- T-S relation
- Level of no motion or elliptic eqn
- Geostrophic balance
- Geostrophic balance
- Linear equation of state
- Hydrostatic balance
The Balance Operator

\[ \delta x = K_b \delta \hat{x} \]

\[
K_b = \begin{pmatrix}
I & 0 & 0 & 0 & 0 \\
K^{ST} & I & 0 & 0 & 0 \\
K_{\zeta T} & K_{\zeta S} & I & 0 & 0 \\
K_{uT} & K_{uS} & K_{u\zeta} & I & 0 \\
K_{vT} & K_{vS} & K_{v\zeta} & 0 & I \\
\end{pmatrix}
\]
The Balance Operator

\( K_{ST} \) from prior (background) \( T-S \) relationship

\[ \delta S_b = \gamma \left. \frac{\partial S}{\partial z} \right|_S \left. \frac{\partial z}{\partial T} \right|_T \delta T \]

\( \gamma = \begin{cases} 0 \\ 1 \end{cases} \) depending on mixed layer
The Balance Operator

\[
K_b = \begin{pmatrix}
I & 0 & 0 & 0 & 0 \\
K_{ST} & I & 0 & 0 & 0 \\
K_{\xi T} & K_{\xi S} & I & 0 & 0 \\
K_{uT} & K_{uS} & K_{u\xi} & I & 0 \\
K_{vT} & K_{vS} & K_{v\xi} & 0 & I
\end{pmatrix}
\]
The Balance Operator

\[
K_{\xi T} = K_{\xi \rho} \left(K_{\rho T} + K_{\rho S} K_{ST}\right) \quad \left\{ \delta \rho = \rho_0 \left(-\alpha \delta T + \beta \delta S\right) \right. \\
K_{\xi S} = K_{\xi \rho} K_{\rho S}
\]

Either:

(i) \( \delta \zeta_b = -\int_{z_r}^{0} \frac{\delta \rho}{\rho_0} dz \) (level of no motion \( z_r \))

(ii) \( \nabla \left( h \nabla \delta \zeta_b \right) = -\nabla \int_{-h}^{0} \int_{z}^{0} \frac{\delta \rho}{\rho_0} dz' \, dz + \ldots \) (define ZETA_ELLIPTIC)
The Balance Operator

\[ K_b = \begin{pmatrix}
I & 0 & 0 & 0 & 0 & 0 \\
K_{ST} & I & 0 & 0 & 0 & 0 \\
K_{\zeta T} & K_{\zeta S} & I & 0 & 0 & 0 \\
K_{uT} & K_{uS} & K_{u\zeta} & I & 0 & 0 \\
K_{vT} & K_{vS} & K_{v\zeta} & 0 & 1 & 0 \\
\end{pmatrix} \]
The Balance Operator

\[
\begin{align*}
K_{uT} &= K_{up} \left( K_{p\rho} + K_{p\zeta} K_{\zeta\rho} \right) \left( K_{\rho T} + K_{\rho S} K_{ST} \right) \\
K_{uS} &= K_{up} \left( K_{p\rho} + K_{p\zeta} K_{\zeta\rho} \right) K_{\rho S} \\
K_{u\zeta} &= K_{up} K_{p\zeta}
\end{align*}
\]

- \( K_{p\rho} \) hydrostatic balance
- \( K_{up} \) geostrophic balance
- \( K_{p\zeta} \) free-surface contribution to \( \rho \)
The Balance Operator

\[ \mathbf{B}_x = \mathbf{K}_b \left( \mathbf{B}_x \right)_u \mathbf{K}_b^T = \begin{pmatrix}
\mathbf{B}_{TT} & \mathbf{B}_{ST}^T & \mathbf{B}_{\zeta T}^T & \mathbf{B}_{uT}^T & \mathbf{B}_{vT}^T \\
\mathbf{B}_{ST} & \mathbf{B}_{SS} & \mathbf{B}_{\zeta S}^T & \mathbf{B}_{uS}^T & \mathbf{B}_{vS}^T \\
\mathbf{B}_{\zeta T} & \mathbf{B}_{\zeta S} & \mathbf{B}_{\zeta \zeta} & \mathbf{B}_{u\zeta}^T & \mathbf{B}_{v\zeta}^T \\
\mathbf{B}_{uT} & \mathbf{B}_{uS} & \mathbf{B}_{u\zeta} & \mathbf{B}_{uu} & \mathbf{B}_{vu}^T \\
\mathbf{B}_{vT} & \mathbf{B}_{vS} & \mathbf{B}_{v\zeta} & \mathbf{B}_{vu} & \mathbf{B}_{vv}
\end{pmatrix} \]
The cross-covariances are computed from a single sea surface height observation using multivariate physical balance relationships.
The cross-covariances are computed from a single temperature observation at the surface using multivariate physical balance relationships.
The cross-covariances are computed from a single U-velocity observation at the surface using multivariate physical balance relationships.
Initial condition prior:
\[ B_x = K_b \Sigma_x C_x \Sigma_x^T K_b^T \]

Surface forcing prior:
\[ B_f = \Sigma_f C_f \Sigma_f^T \quad \text{No balance} \]

Open boundary condition prior:
\[ B_b = \Sigma_b C_b \Sigma_b^T \quad \text{No balance} \]

Model error prior:
\[ Q = K_b \Sigma_q C_q \Sigma_q^T K_b^T \]
Preconditioning Again

General form of the *prior* error covariance matrix:

\[ D = K_b \Sigma C \Sigma^T K_b^T \]

Introduce a new variable:

\[ v = U^{-1} \delta z \]

where

\[ D = UU^T \]

\[ U = K_b \Sigma C^{1/2} \]
The Conjugate Gradient Algorithm

cgradient.h in v-space to minimize $J(v)$

\[
\hat{v}_k = v_k + \tau_k h_k \\
\hat{g}_k = C^{T/2} \Sigma^T K_b^T \partial J / \partial \delta \hat{z}_k \\
\alpha_k = -\tau_k h_k^T g_k / \left( h_k^T (\hat{g}_k - g_k) \right) \\
v_{k+1} = v_k + \alpha_k h_k
\]

\[
g_{k+1} = g_k + (\alpha_k / \tau_k) (\hat{g}_k - g_k) \\
\beta_{k+1} = g_{k+1}^T g_{k+1} / g_k^T g_k \\
h_{k+1} = -g_{k+1} + \beta_{k+1} h_k \\
\delta z_{k+1} = K_b \Sigma C^{1/2} v_{k+1}
\]

trial step
gradient @ trial step
optimum step
new starting point
gradient @ new point
new descent direction
project into state-space
The Lanczos Connection

Gain (primal form):

\[ K = K_b \Sigma C^{1/2} (I + D^{T/2}G^T R^{-1}G D^{1/2})^{-1} C^{T/2} \Sigma^T K_b G^T R^{-1} \]

Practical gain matrix:

\[ \tilde{K} = K_b \Sigma C^{1/2} V_k T_k^{-1} V_k^T C^{T/2} \Sigma^T K_b G^T R^{-1} \]

Useful for diagnostic applications (Lecture 5)
(The Lanczos vectors are in ADJname)
Background Quality Control
(define BGQC)

• Some observations will be outliers for a variety of reasons (e.g. bad obs, bad model, or both, non-Gaussian behavior, etc)
• It is important to exclude these data from the data assimilation system since they can adversely affect the analysis.
• Observations are screened in ROMS according to the background error and observation error variances.
• The approach used in ROMS parallels that used in the ECMWF NWP system.
• An observation is rejected if the normalized innovation exceeds a specified multiple of the standard expected error.
• Specifically:

\[
\left( y_i - H_i(x_b) \right)^2 > \alpha^2 \left( 1 + \sigma_o^2 / \sigma_b^2 \right)
\]

• \( \alpha \) is a user-specified parameter.
(see Andersson and Järvinen (1999, QJRMS, 125, 697-722).
Background Quality Control
(define BGQC)

Frequency distribution, $f$, (i.e. pdf) of 4D-Var innovations.

Distribution of $\hat{f} = \sqrt{-2\ln(f/\max(f))}$. The red lines show $y = \pm \left|x/(\sigma_b^2 + \sigma_o^2)^{1/2}\right|$. Choose $\alpha$ based on red line.
Issues & Things to do

- Relax horizontal homogeneity and isotropy of $L_x$ and $L_y$ correlation lengths.
- Elliptic solver for free-surface balance:
  - cannot handle islands at the moment
  - add additional boundary condition option
- Cannot assimilate obs right at the open boundary.
- Div and curl of $\delta \tau$ are not constrained.
- No restart option for 4D-Var.
- Variational bias correction.
- Variational QC.
Summary

• Lanczos formulation of CG: cgradient.h
• Lanczos vectors saved in ADJname
• Covariance models using diffusion operators:
  define VCONVOLUTION
  define IMPLICIT_VCONV, etc
  $\Sigma$ - tl_variability.F
  $\Sigma^T$ - ad_variability.F
  $C^{1/2}$ - tl_convolution.F
  $C^{T/2}$ - ad_convolution.F
• Multivariate balance operator:
  define BALANCE_OPERATOR
  $K_b$ - tl_balance.F
  $K^T_b$ - ad_balance.F
• Background QC: define BGQC
 References


