The Oneshot Optimization Strategy: Simultaneous Model Spin-Up and Parameter Optimization

Claudia Kratzenstein, T. Slawig, J. Piwonski

Christian-Albrechts-University Kiel, Germany, Cluster of Excellence "The Future Ocean"

joint work with A. Griewank et. al. (HU Berlin) and N. Gauger et. al. (RWTH Aachen)

July 1, 2014, Kiel
1 Motivation

2 The Oneshot Optimization Strategy

3 Numerical Results
Motivation

Fit model output to observed data
Optimize parameters $u$ in the underlying model

Formulation of the mathematical problem

$y \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $m \ll n$

$$
\min J(y, u) := \frac{1}{2} \| y - y_{data} \|^2 + \frac{\alpha}{2} \| u - u_{est} \|^2
$$

s.t. model equations are fulfilled, i.e. $0 = e(y, u)$. 
Oneshot Optimization Strategy

**Problem:**

\[
\min_{y,u} J(y, u) \quad s.t. \quad e(y, u) = 0
\]

**Idea:** Use formulation with a fixed point equation with contraction \(G:\)

\[
y = G(y, u) \iff e(y, u) = 0
\]

**Lagrangian:** \(L(y, \tilde{y}, u) = J(y, u) + \tilde{y}^T(G(y, u) - y)\)

1st order necessary optimality conditions at \((y^*, \tilde{y}^*, u^*)\):

\[
0 = \frac{\partial L}{\partial y} = J_y(y^*, u^*) + \tilde{y}^*^T G_y(y^*, u^*) - \tilde{y}^*^T
\]

\[
0 = \frac{\partial L}{\partial \tilde{y}} = G(y^*, u^*) - y^*
\]

\[
0 = \frac{\partial L}{\partial u} = J_u(y^*, u^*) + \tilde{y}^*^T G_u(y^*, u^*)
\]

\(\Rightarrow\) **Coupled Oneshot iteration:**

\[
y_{k+1} = G(y_k, u_k) \quad \text{feasibility}
\]

\[
\tilde{y}_{k+1}^T = \tilde{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k) \quad \text{adjoint feasibility}
\]

\[
u_{k+1} = u_k - B_k^{-1} \left( J_u(y_k, u_k) + \tilde{y}_k^T G_u(y_k, u_k) \right) \quad \text{optimality}
\]
Problem:
\[
\min_{y,u} J(y,u) \quad s.t. \quad e(y,u) = 0
\]

Idea: Use formulation with a fixed point equation with contraction \(G\):
\[
y = G(y,u) \iff e(y,u) = 0
\]

Lagrangian: \(L(y, \bar{y}, u) = J(y,u) + \bar{y}^T (G(y,u) - y)\)

1st order necessary optimality conditions at \((y^*, \bar{y}^*, u^*)\):
\[
0 = \frac{\partial L}{\partial y} = J_y(y^*, u^*) + \bar{y}^*^T G_y(y^*, u^*) - \bar{y}^*^T \\
0 = \frac{\partial L}{\partial \bar{y}} = G(y^*, u^*) - y^* \\
0 = \frac{\partial L}{\partial u} = J_u(y^*, u^*) + \bar{y}^*^T G_u(y^*, u^*)
\]

⇒ Coupled Oneshot iteration:
\[
\begin{align*}
y_{k+1} &= G(y_k, u_k) & \text{feasibility} \\
\bar{y}_{k+1}^T &= \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k) & \text{adjoint feasibility} \\
u_{k+1} &= u_k - B_k^{-1} (J_u(y_k, u_k) + \bar{y}_k^T G_u(y_k, u_k)) & \text{optimality}
\end{align*}
\]
Choice of Preconditioner B

Definition (doubly augmented Lagrangian):

\[
L^a(y, \bar{y}, u) = L(y, \bar{y}, u) + \frac{\alpha}{2} \|G(y, u) - y\|^2 + \frac{\beta}{2} \|J_y(y, u) + \bar{y}^\top G_y(y, u) - \bar{y}\|^2
\]

▷ [Griewank, Hamdi 2008]:

\[
B = \frac{1}{\sigma} \left( \alpha G_u^\top G_u + \beta L_{yu}^\top L_{yu} + L_{uu} \right) \approx \frac{1}{\sigma} \nabla_{uu}^2 L^a
\]

where \(\sigma, \alpha\) and \(\beta\) are carefully chosen weights.

▷ Calculation of derivative information using Automatic/Algorithmic Differentiation (AD).
Choice of Preconditioner B

Definition (doubly augmented Lagrangian):

\[ L^a(y, \bar{y}, u) = L(y, \bar{y}, u) + \frac{\alpha}{2} \| G(y, u) - y \|^2 + \frac{\beta}{2} \| J_y(y, u) + \bar{y}^\top G_y(y, u) - \bar{y} \|^2 \]

▷ [Griewank, Hamdi 2008]:

\[
B = \frac{1}{\sigma} (\alpha G_u^\top G_u + \beta L_{yu}^\top L_{yu} + L_{uu}) \approx \frac{1}{\sigma} \nabla^2_{uu} L^a
\]

where \( \sigma, \alpha \) and \( \beta \) are carefully chosen weights.

▷ Calculation of derivative information using Automatic/Algorithmic Differentiation (AD).
Automatic Differentiation

Example: \( y_1 = F(x_1, x_2) = [\sin(x_1/x_2) + x_1/x_2 - \exp(x_2)] \times [x_1/x_2 - \exp(x_2)] \)

The **main ingredient**: application of the chain rule

\[ \nabla_x h(g(x)) = \sum_{i=1}^{k} \frac{\partial h}{\partial g_i} \nabla g_i(x), \]

where \( g : \mathbb{R}^n \to \mathbb{R}^k \) and \( h : \mathbb{R}^k \to \mathbb{R}^m \).

<table>
<thead>
<tr>
<th>Mode</th>
<th>add. Input</th>
<th>Output</th>
<th>Costs</th>
<th>Number of calls for full Jacobian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>( \dot{x} \in \mathbb{R}^n )</td>
<td>( F'(x)\dot{x} )</td>
<td>( k \times \text{costs}(F(x)), k \in [2, \frac{5}{2}] )</td>
<td>( n )</td>
</tr>
<tr>
<td>Reverse</td>
<td>( \bar{y} \in \mathbb{R}^m )</td>
<td>( \bar{y}^\top F'(x) )</td>
<td>( k \times \text{costs}(F(x)), k \in [3, 4] )</td>
<td>( m )</td>
</tr>
</tbody>
</table>
Tools:

(see www.autodiff.org)

- ADOL-C (C, C++)
- TAPENADE (C/C++, Fortran77, Fortran95)
- ADiMat (MATLAB)
- ADMAT / ADMIT (MATLAB)
- TAF (Fortran77, Fortran95)
- TAMC (Fortran77)
- TAC++ (C++)
1 Motivation

2 The Oneshot Optimization Strategy

3 Numerical Results
Tracer transport equations for phosphate $y_1$ and dissolved organic phosphorus $y_2$ in the euphotic zone $\Omega_1$ and noneuphotic zone $\Omega_2$

- use pre-computed transport operators (Khatiwala et.al, 2005)
- 6 (7) parameters to optimize

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Unit</th>
<th>$u_{opt}$</th>
<th>$u_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>$\lambda$</td>
<td>remineralization rate of DOP</td>
<td>$1/d$</td>
<td>0.5</td>
</tr>
<tr>
<td>$u_2$</td>
<td>$\alpha$</td>
<td>maximum community production rate</td>
<td>$1/d$</td>
<td>2.0</td>
</tr>
<tr>
<td>$u_3$</td>
<td>$\sigma$</td>
<td>fraction of DOP, $\bar{\sigma} = (1 - \sigma)$</td>
<td>-</td>
<td>0.67</td>
</tr>
<tr>
<td>$u_4$</td>
<td>$K_N$</td>
<td>half saturation constant of N</td>
<td>$mmolP/m^3$</td>
<td>0.5</td>
</tr>
<tr>
<td>$u_5$</td>
<td>$K_l$</td>
<td>half saturation constant of light</td>
<td>$W/m^2$</td>
<td>30.0</td>
</tr>
<tr>
<td>$u_6$</td>
<td>$K_{H2O}$</td>
<td>attenuation of water</td>
<td>$1/m$</td>
<td>0.02</td>
</tr>
<tr>
<td>$u_7$</td>
<td>$b$</td>
<td>sinking velocity exponent</td>
<td>-</td>
<td>0.858</td>
</tr>
</tbody>
</table>

**Discretization:**

- $dim y_i = 52749, i = 1, 2$
- 2880 steps per year
- 1 model spin-up with fixed parameters takes $> 5000$ model years

- synthetic data, computed by this model
- real world data from World Ocean Atlas 2009
Derivatives computed from only one model year of the update of \( y \)

- TAF only applied on bio-chemistry \( q \)

\[
y_{k,j+1} = A_{imp,j}(A_{exp,j}y_{k,j} + q_j(y_{k,j}, u)), \quad j = 0, \ldots, 2880 - 1
\]

\[
y_{k+1} = y_{k,2880}
\]

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Mode of Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_y )</td>
<td>analytically</td>
</tr>
<tr>
<td>( J_u )</td>
<td>analytically</td>
</tr>
<tr>
<td>( \bar{y}^\top G_y, \bar{y}^\top G_u )</td>
<td>one reverse sweep of AD + analytically for linear parts</td>
</tr>
<tr>
<td>( G_u )</td>
<td>finite differences</td>
</tr>
<tr>
<td>( J_{yu} = 0 )</td>
<td>analytically</td>
</tr>
<tr>
<td>( \bar{y}^\top G_{yu}, \bar{y}^\top G_{uu} )</td>
<td>finite differences after AD computation of ( \bar{y}^\top G_y, \bar{y}^\top G_u )</td>
</tr>
</tbody>
</table>
Numerical Results Synthetic Data

\[
\min J(y, u) := \frac{1}{2} \|y - y_{data}\|^2 + \frac{\alpha}{2} \|u - u_{est}\|^2 \quad \text{s.t.} \quad y = G(y, u)
\]

Reduction of the cost functional for different \(\alpha\).
Numerical Results Synthetic Data

\[ \| y - y_{\text{data}} \| \]

\[ \| u - u_{\text{est}} \| \] for different weighting factors \( \alpha \)

- \( \alpha = 100 \)
- \( \alpha = 1 \)
- \( \alpha = 0 \)
Numerical Results Synthetic Data

Difference to data, PO4 uppermost layer, $\alpha = 0$
Numerical Results Synthetic Data

Difference to data, PO4 fifth layer b.s.l. (550m), $\alpha = 0$
Numerical Results Synthetic Data

Parameter values during the optimization for different $\alpha$. 

Claudia Kratzenstein The Oneshot Optimization Strategy 15 / 27
Numerical Results Synthetic Data

\[ J(y, u) := \frac{1}{2} \| y - y_{data} \|^2 + \frac{\alpha}{2} \| u - u_{est} \|^2 \]

Results of the optimization for \( \alpha = 0.01 \), \( u_{est} \neq u_{opt} \),
\( u_{est} = (0.45, 3.0, 0.6, 0.8, 27.0, 0.22) \).
Numerical Results Synthetic Data

Parameter values during the optimization for $\alpha = 0.01$, $u_{est} \neq u_{opt}$, $u_{est} = (0.45, 3.0, 0.6, 0.8, 27.0, 0.22)$. 
## Results on Computational Time

<table>
<thead>
<tr>
<th></th>
<th>Spin-up</th>
<th>Spin-up and adjoint state</th>
<th>Oneshot</th>
</tr>
</thead>
<tbody>
<tr>
<td>elapsed time of 1 step factor</td>
<td>10.9 s</td>
<td>43.4 s</td>
<td>292.4 s</td>
</tr>
<tr>
<td>elapsed time of 500 steps factor</td>
<td>1:27:31 h</td>
<td>5:41:35 h</td>
<td>39:56:07 h</td>
</tr>
<tr>
<td>elapsed time until acceptance factor</td>
<td>30:13:07 h</td>
<td>113:15:12 h</td>
<td>409:34:42 h</td>
</tr>
</tbody>
</table>

Initialize $u_0, y_0, \bar{y}_0$

**DO k=1:500 (Pre-Iterations)**

$y_{k+1} = G(y_k, u_k)$

$\bar{y}_{k+1} = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k)$

$u_{k+1} = u_k$

**END DO**

**DO k=501:1000 (exact Oneshot steps)**

$y_{k+1} = G(y_k, u_k)$

$\bar{y}_{k+1} = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k)$

Compute $B$

$u_{k+1} = u_k - B^{-1} (L_u(y_k, \bar{y}_k, u_k))$

**END DO**

**DO until convergence (inexact Oneshot step)**

$y_{k+1} = G(y_k, u_k)$

$\bar{y}_{k+1} = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k)$

Compute $B$ every 5 steps

$u_{k+1} = u_k - B^{-1} (L_u(y_k, \bar{y}_k, u_k))$

$k = k+1$

**END DO**
Results on Computational Time

<table>
<thead>
<tr>
<th></th>
<th>Spin-up</th>
<th>Spin-up and adjoint state</th>
<th>Oneshot</th>
</tr>
</thead>
<tbody>
<tr>
<td>elapsed time of 1 step factor</td>
<td>10.9 s</td>
<td>43.4 s</td>
<td>292.4 s</td>
</tr>
<tr>
<td>elapsed time of 500 steps factor</td>
<td>1:27:31 h</td>
<td>5:41:35 h</td>
<td>39:56:07 h</td>
</tr>
<tr>
<td>elapsed time until acceptance factor</td>
<td>30:13:07 h</td>
<td>113:15:12 h</td>
<td>409:34:42 h</td>
</tr>
</tbody>
</table>

Initialize $u_0, y_0, \bar{y}_0$

DO k=1:500 (Pre-iterations)

\[
y_{k+1} = G(y_k, u_k)
\]
\[
\bar{y}_{k+1} = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k)
\]
\[
u_{k+1} = u_k
\]

END DO

DO k=501:1000 (exact Oneshot steps)

\[
y_{k+1} = G(y_k, u_k)
\]
\[
\bar{y}_{k+1} = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k)
\]

Compute $B$
\[
u_{k+1} = u_k - B^{-1}(L_u(y_k, \bar{y}_k, u_k))
\]

END DO

DO until convergence (inexact Oneshot step)

\[
y_{k+1} = G(y_k, u_k)
\]
\[
\bar{y}_{k+1} = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k)
\]

Compute $B$ every 5 steps
\[
u_{k+1} = u_k - B^{-1}(L_u(y_k, \bar{y}_k, u_k))
\]
\[k = k+1\]

END DO
Results of the optimization for $\alpha = 0.01$, $u_{est} \neq u_{opt}$, 
$u_{est} = (0.45, 3.0, 0.6, 0.8, 27.0, 0.22)$. 
$J(y, u) := \frac{1}{2} \|y - y_{data}\|^2 + \frac{\alpha}{2} \|u - u_{est}\|^2, \ u_{est} \neq u_{opt}$
Numerical Results Real World Data ($y_{data}$ from WOA 2009)
Numerical Results Real World Data ($y_{data}$ from WOA 2009)

Difference to data after 15,000 Oneshot steps, uppermost layer
Numerical Results Real World Data ($y_{data}$ from WOA 2009)

Difference to data after 15,000 Oneshot steps, fifth layer b.s.l (550m)
Parameter values during the optimization process.
Numerical Results Synthetic Data, 7 Parameters

\[ \alpha = 100, 7 \text{ parameters} \]

Cost functional \( J \)

\[ \| u - u_{\text{est}} \| \]

\[ \| y - y_{\text{data}} \| \]

\( u_2 \)

\( u_6 \)

\( u_7 \)
Numerical Results Synthetic Data, 7 Parameters

![Graph showing the results of the optimization process for 7 parameters. The graph displays the cost function and the mean value for the norm of the solution (normu) and the norm of the residual (normy). The solutions are labeled with subscripts ranging from 1 to 7. The x-axis represents different values, and the y-axis shows the corresponding values for the solutions and the cost function. The graph illustrates the optimization process and convergence of the solutions towards the optimal values.]
Summary and Outlook

Oneshot Method:

\[ y_{k+1} = G(y_k, u_k) \]  Feasibility

\[ \bar{y}_{k+1}^T = \bar{y}_k^T G_y(y_k, u_k) + J_y(y_k, u_k) \]  Adjoint Feasibility

\[ u_{k+1} = u_k - B_k^{-1} \left( J_u(y_k, u_k) + \bar{y}_k^T G_u(y_k, u_k) \right) \]  Optimality


▷ Application to a 3D biogeochemical ocean model with synthetic and WOA data (K., *My Dissertation*, (2014))