The Collocation Solution of BVP for a System of ODEs in Photosynthetic Carbon Metabolism Ultimately for Parameter Estimation

EUN CHANG
VIRGINIA TECH

FEBRUARY 20, 2010

2010 SIAM STUDENT CONFERENCE

CHANGER@VT.EDU

Collaborators

- **Advisor: Dr. Eric de Sturler**

Department of Mathematics

Virginia Tech

sturler@vt.edu

- **Dr. Xin-guang Zhu**

Principle Investigator

Chinese Academy of Sciences

Research Scientist

Department of Physiological Molecular Plant Biology

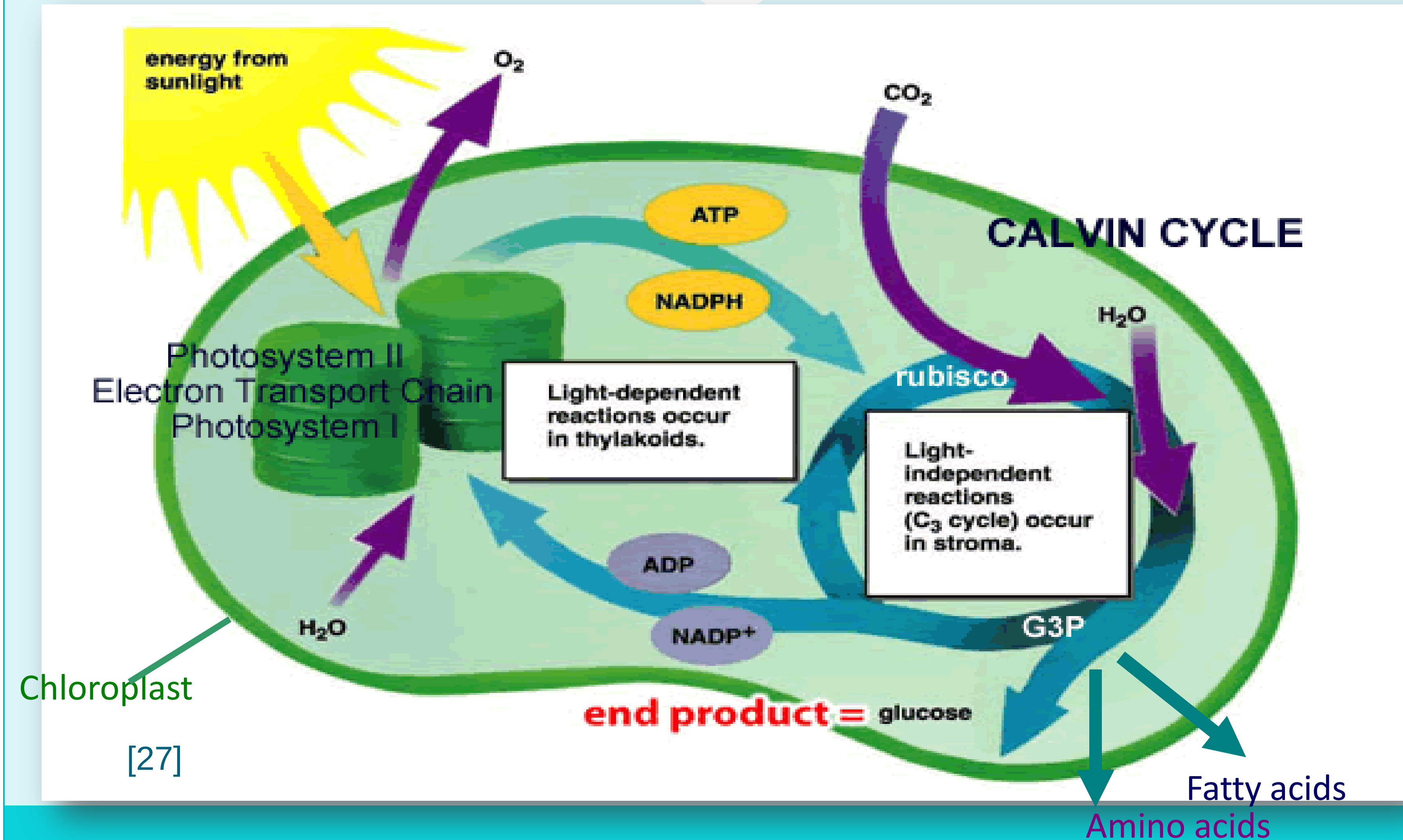
University of Illinois at Urbana-Champaign

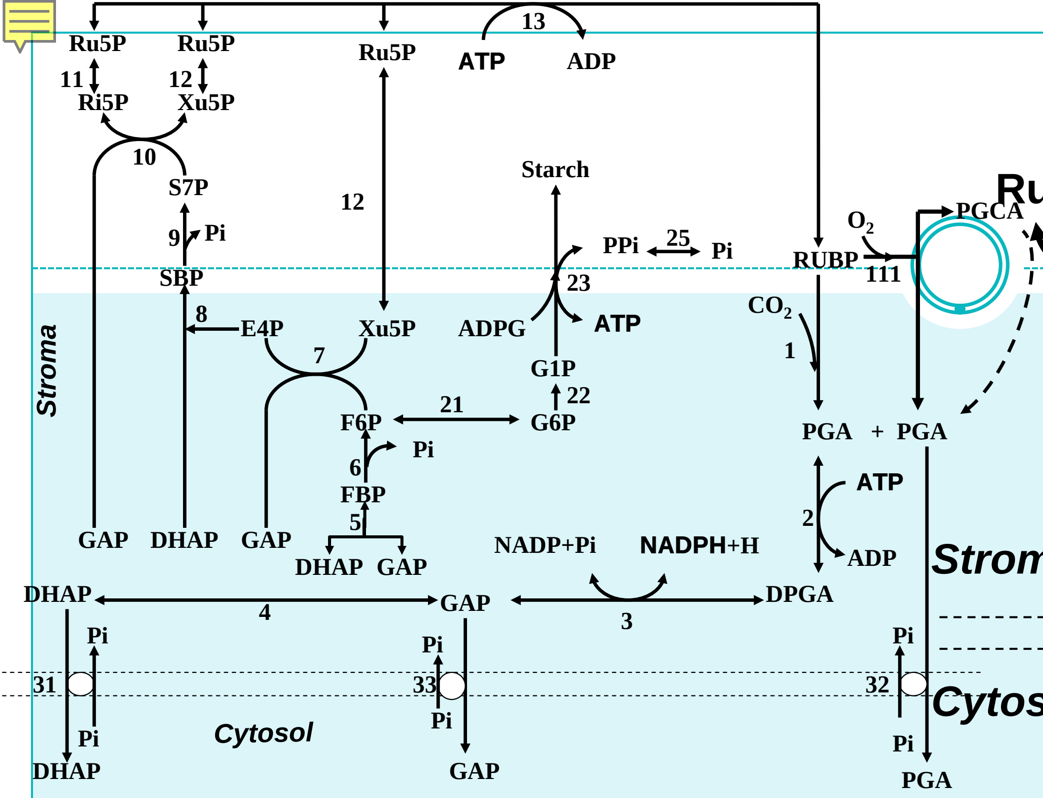
Zhu3@uiuc.edu

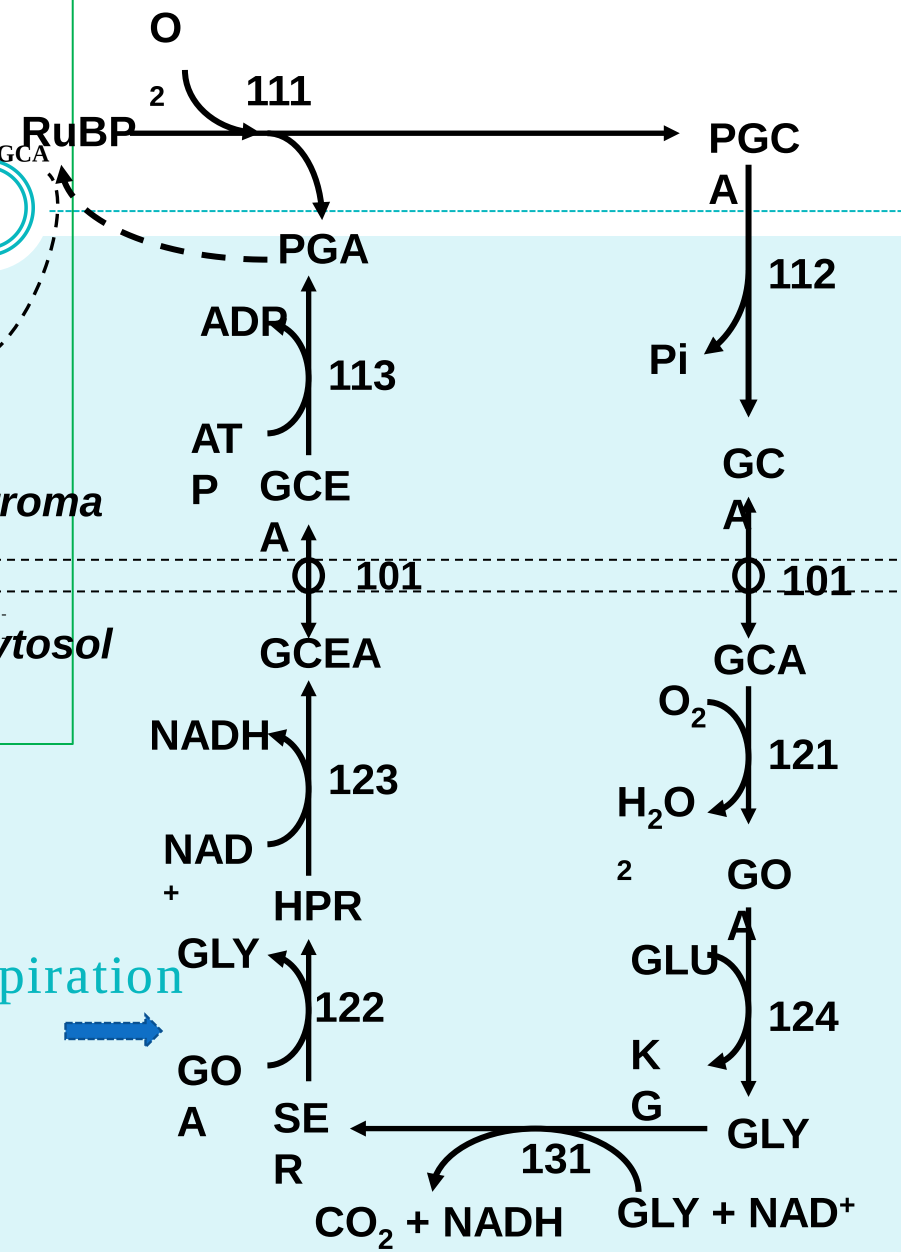
WHY PHOTOSYNTHETIC CARBON METABOLISM?

1. THE MOST IMPORTANT BIOLOGICAL PROCESS ON EARTH
   - SURVIVE! Food, hospitable environment & Energy
     - Generate Carbohydrate and oxygen.
       - Photosynthesis Output of Carbohydrate: 160 billion metric tons yearly. = 352000000000 pounds.
2. Photorespiration: Not understood completely[26]
3. Extended photosynthetic carbon metabolism[20]
   - No current dynamic model of photosynthetic carbon metabolism includes all the reactions in the following.
     1. Calvin cycle
     2. Photorespiratory metabolism
     3. Starch synthesis
     4. Suc synthesis

K-12 Photosynthesis: Light Reactions and the Calvin Cycle.







Calvin cycle, starch synthesis, and triose phosphate.

Photorespiration

[21]

WHAT PARAMETERS?

- **Enzymes for improved crop yield Or increased photosynthetic rate [de Sturler, Zhu, Long]**
  1. Biotechnology can engineer each enzyme to alter resource allocation between the enzymes
  2. 38 enzymes involved in extended photosynthetic carbon metabolism
  3. Given a fixed resource of total protein-nitrogen, $\mathbf{10^9}$ testing choices for optimal distribution of protein nitrogen into different enzymes.

     ⏭ **WHICH ENZYMES?** ⏭**NEED TO IDENTIFY**

- **The Maximum Rate of some reactions or of all Enzymes in the Photosynthetic Carbon Metabolism**
- 7 Proteins that are not necessary for photosynthesis but are required for photorespiration
- Initial Conditions
- Kinetic Constants Used in Rate Equations of each Metabolite[21]

PHOTOSYNTHESIS AND ENERGY

- **<u>2010 The State of the Union</u>**
  - **"<u>...Continued investment in advanced biofuel...</u>"**
  - **Increase Crop Yield →More Biofuel**
  - Ethanol Fuel: Fermented sugars from sugar cane, wheat, corn (C_4 plants) etc.
  - E10: Mandatory in 10 States.

  - E85: is used in Sweden. 7 stations in VA (2 open to public), normally 30% cost less than gasoline.
  - E100 is widely used in Brazil.

[28,30]

A SYSTEM OF ODES.

$$\frac{d[RuBP]}{dt} = v_{13} - v_1 - v_{111};$$

$$\frac{d[PGA]}{dt} = 2v_1 - v_2 - v_{32} + v_{111} + v_{113};$$

$$\frac{d[DPGA]}{dt} = v_2 - v_3;$$

$$\frac{d[T3P]}{dt} = v_3 - 2v_5 - v_7 - v_8 - v_{10} - v_{31} - v_{33};$$

$$\frac{d[FBP]}{dt} = v_5 - v_6;$$

$$\frac{d[E4P]}{dt} = v_7 - v_8;$$

$$\frac{d[S7P]}{dt} = v_9 - v_{10};$$

$$\frac{d[SBP]}{dt} = v_8 - v_9;$$

$$\frac{d[ATP]}{dt} = v_{16} - v_2 - v_{23} - v_{13} - v_{113};$$

$$\frac{d[HexP]}{dt} = v_6 - v_7 - v_{23};$$

$$\frac{d[PenP]}{dt} = v_7 + 2v_{10} - v_{13};$$

$$\frac{d[GOAc]}{dt} = V_{121} - V_{122} - V_{124}$$

$$\frac{d[SERc]}{dt} = v_{131} - v_{122}$$

$$\frac{d[GLYc]}{dt} = v_{122} + v_{124} - 2v_{131}$$

$$\frac{d[HPRc]}{dt} = v_{122} - v_{123}$$

$$\frac{d[GCEAc]}{dt} = v_{123} - v_{1in} + v_{1out}$$

$$\frac{d[GCEA]}{dt} = v_{1in} - v_{113} - v_{1out}$$

$$\frac{d[GCA]}{dt} = v_{112} - v_{2out} + v_{2in}$$

$$\frac{d[PGCA]}{dt} = v_{111} - v_{112}$$

$$\frac{d[GCAc]}{dt} = v_{2out} - v_{121} - v_{2in}$$

Where

$$v_{13} = \frac{V_{13}\left(ATP \times Ru5P\right)}{\left(ATP\left(1+\frac{ADP}{K_{I134}}\right)+K_{M132}\left(1+\frac{ADP}{K_{I135}}\right)\right)\times\left(Ru5P+K_{M131}\left(1+\frac{GAP}{K_{I131}}+\frac{RuBP}{K_{I132}}+\frac{Pi}{K_{I133}}\right)\right)}$$

$$\vdots$$

$$v_{1} = \frac{RuBP \times W_C \times \min\left(1, \frac{RuBP}{E_t}\right)}{\left(RuBP+K_r\left(1+\frac{PGA}{K_{I11}}+\frac{FBP}{K_{I12}}+\frac{SBP}{K_{I13}}+\frac{Pi}{K_{I14}}+\frac{NADPH}{K_{I15}}\right)\right)} \; ; \quad W_C = \frac{V_{C\max} \times CO_2}{CO_2 + K_{M11}\left(1+\frac{O_2}{K_{M12}}\right)}$$

$$v_{111} = \frac{RuBP \times W_O \times \min\left(1, \frac{RuBP}{E_t}\right)}{\left(RuBP+K_r\left(1+\frac{PGA}{K_{I11}}+\frac{FBP}{K_{I12}}+\frac{SBP}{K_{I13}}+\frac{PI}{K_{I14}}+\frac{NADPH}{K_{I15}}\right)\right)} \; ; \quad W_O = \frac{V_{111} \times O_2}{O_2 + K_O\left(1+\frac{CO_2}{K_C}\right)}$$

$$\vdots$$

Stiffness of ODEs

- Large negative real part of eigenvalues along with others normal magnitude.
- Local Jacobian , $\frac{\partial f}{\partial y}$, has large eigenvalues.
- Need A-Stability Solver or extremely small time steps.
  - IRK_s (Better Stability with the higher order)= Collocation Method
    - Simple One-Step scheme
- Note that ODE stiff is not related stiffness in parameters as in “Stiff” parameter & “Sloppy” parameter in Gutenkunst[7] and Ashyraliyev[3]

Implicit S-Stage Runge-Kutta (IRK)

$$y_{i+1} = y_i + h_i \sum_{s=1}^{S} \beta_s f_{is}; \quad 1 \le i \le K.$$

$$f_{is} = f\left(t_{is},\ y_i + h_i \sum_{l=1}^{S} \alpha_{sl} f_{il} \right).\ 1 \le s \le S$$

$$t_{is} = t_i + h_i \rho_s; \qquad 1 \le s \le S, \qquad 1 \le i \le K$$

$$f_{is} = f\left(t_{is}, y_{is}\right);\ y_{is} := y_i + h_i \sum_{l=1}^{S} \alpha_{sl} f_{il} = \text{appr of } y(t_{is})$$

with $0 \le \rho_1 < \rho_2 < \cdots < \rho_S \le 1.$

The points ρ_s are distinct.

IRK_s := This special case of IRK

[1,10]

Construction of Weight For IRK_s

where aggregate α and β are

$$\sum_{l=1}^{S} \alpha_{sl} = \rho_s . \text{And} \sum_{l=1}^{S} \beta_l = 1. \quad 1 \le s \le S$$

The Butcher Diagram.

ρ_1	α_{11}	$\cdots$	α_{1S}
$\vdots$	$\vdots$		$\vdots$
ρ_S	α_{S1}	$\cdots$	α_{SS}
	β_1	$\cdots$	β_S

$\Rightarrow$

C	A
	B^T

Collocation Method : ODE Solver

1. Discretization
   - Time domain $t_0 < t_1 < \ldots < t_k < \ldots < t_N, \quad I_k = [t_{k-1}, t_k].$
2. Choose Basis Function:
   - Lagrange Polynomial (for each metabolite and each interval)

$$\tilde{y}_{k,j}(t) = f(l_k)\frac{(t-m_k)(t-r_k)}{(l_k-m_k)(l_k-r_k)} + f(m_k)\frac{(t-l_k)(t-r_k)}{(m_k-l_k)(m_k-r_k)}$$

$$+ f(r_k)\frac{(t-l_k)(t-m_k)}{(r_k-l_k)(r_k-m_k)}$$

$$:= A_{k,j}\frac{(t-m_k)(t-r_k)}{(l_k-m_k)(l_k-r_k)} + B_{k,j}\frac{(t-l_k)(t-r_k)}{(m_k-l_k)(m_k-r_k)} + C_{k,j}\frac{(t-l_k)(t-m_k)}{(r_k-l_k)(r_k-m_k)}$$

Where $A_{k,j} = f(l_k), B_{k,j} = f(m_k)$ & $C_{k,j} = f(r_k)$, l, m, r are timesteps, and $f(\bullet)$ is a function value at the time.

$A_{k,j}, B_{k,j}, C_{k,j}$ are unknowns $\rightarrow 3 \cdot K \cdot J$ unknowns

Need $3 \cdot K \cdot J$ Equations

1. Left Continuity Condition: $K \cdot J$
2. Right Differentiability Condition: $K \cdot J$
3. Collocation Condition: $K \cdot J$
   - Satisfy the given model

- → Highly Nonlinear Equations
- → Need Jacobian

→ $3 \cdot K \cdot J$ algebraic equations

Left Continuity Condition

$$\tilde{y}_{k-1,j}(l_k)-\tilde{y}_{k,j}(l_k)=A_{k-1,j}\frac{(l_k-m_{k-1})(l_k-r_{k-1})}{(l_{k-1}-m_{k-1})(l_{k-1}-r_{k-1})}+B_{k-1,j}\frac{(l_k-l_{k-1})(l_k-r_{k-1})}{(m_{k-1}-l_{k-1})(m_{k-1}-r_{k-1})}$$

$$+C_{k-1,j}\frac{(l_k-l_{k-1})(l_k-m_{k-1})}{(r_{k-1}-l_{k-1})(r_{k-1}-m_{k-1})}-A_{k,j}\frac{(l_k-m_k)(l_k-r_k)}{(l_k-m_k)(l_k-r_k)}$$

$$-B_{k,j}\frac{(l_k-l_k)(l_k-r_k)}{(m_k-l_k)(m_k-r_k)}-C_{k,j}\frac{(l_k-l_k)(l_k-m_k)}{(r_k-l_k)(r_k-m_k)}=0.$$

Note that $l_k=r_{k-1}$

$$\Rightarrow \tilde{y}_{k-1,j}(l_k)-\tilde{y}_{k,j}(l_k)=C_{k-1,j}-A_{k,j}$$

$\Rightarrow (K-1)J$ equations

Right Differentiability Condition

$\Rightarrow (K-1)J$ equations

$$\tilde{y}'_{k,j}(r_k) - \tilde{y}'_{k+1,j}(r_k) = A_{k,j}\frac{(r_k - m_k) + (r_k - r_k)}{(l_k - m_k)(l_k - r_k)} + B_{k,j}\frac{(r_k - l_k) + (r_k - r_k)}{(m_k - l_k)(m_k - r_k)}$$

$$+C_{k,j}\frac{(r_k - l_k) + (r_k - m_k)}{(r_k - l_k)(r_k - m_k)} - A_{k+1,j}\frac{(r_k - m_{k+1}) + (r_k - r_{k+1})}{(l_{k+1} - m_{k+1})(l_{k+1} - r_{k+1})}$$

$$-B_{k+1,j}\frac{(r_k - l_{k+1}) + (r_k - r_{k+1})}{(m_{k+1} - l_{k+1})(m_{k+1} - r_{k+1})} - C_{k+1,j}\frac{(r_k - l_{k+1}) + (r_k - m_{k+1})}{(r_{k+1} - l_{k+1})(r_{k+1} - m_{k+1})} = 0.$$

Since $r_k = l_{k+1}$ and let $h_k = r_k - l_k$

$$\Rightarrow \tilde{y}'_{k,j}(r_k) - \tilde{y}'_{k+1,j}(r_k) = \frac{A_{k,j} - 4B_{k,j} + 3C_{k,j}}{h_k} + \frac{3A_{k+1,j} - 4B_{k+1,j} + C_{k,j}}{h_{k+1}}$$

Boundary Conditions

$\Rightarrow 2J$ equations

- **LBC**

$$\tilde{y}_{1,j}(l_1) - IC = A_{1,j}\frac{(l_1 - m_1)(l_1 - r_1)}{(l_1 - m_1)(l_1 - r_1)} + B_{1,j}\frac{(l_1 - l_1)(l_1 - r_1)}{(m_1 - l_1)(m_1 - r_1)}$$

$$+C_{1,j}\frac{(l_1 - l_1)(l_1 - m_1)}{(r_1 - l_1)(r_1 - m_1)} - IC_j$$

$$= A_{1,j} - IC_j = 0 \text{ where } IC \text{ represents initial conditions.}$$

- **RBC**

$$\tilde{y}_{k,j}(r_k) - EC_j = A_{k,j}\frac{(r_k - m_k)(r_k - r_k)}{(l_k - m_k)(l_k - r_k)} + B_{k,j}\frac{(r_k - l_k)(r_k - r_k)}{(m_k - l_k)(m_k - r_k)}$$

$$+C_{k,j}\frac{(r_k - l_k)(r_k - m_k)}{(r_k - l_k)(r_k - m_k)} - EC_j$$

$$= C_{k,j} - EC_j = 0 \text{ where } EC \text{ represents end conditions.}$$

Collocation Condition

$\Rightarrow KJ$ equations

$$\tilde{y}'_{k,j}(m_k) = A_{k,j}\frac{(m_k - m_k)+(m_k - r_k)}{(l_k - m_k)(l_k - r_k)} + B_{k,j}\frac{(m_k - l_k)+(m_k - r_k)}{(m_k - l_k)(m_k - r_k)}$$

$$+C_{k,j}\frac{(m_k - l_k)+(m_k - m_k)}{(r_k - l_k)(r_k - m_k)} = f\left(m_k, \tilde{y}(m_k)\right)$$

$$\Rightarrow \dot{\tilde{y}}_{k,j}(m_k) - f\left(m_k, \tilde{y}(m_k)\right) = 0$$

For example: (Equidistance Case)

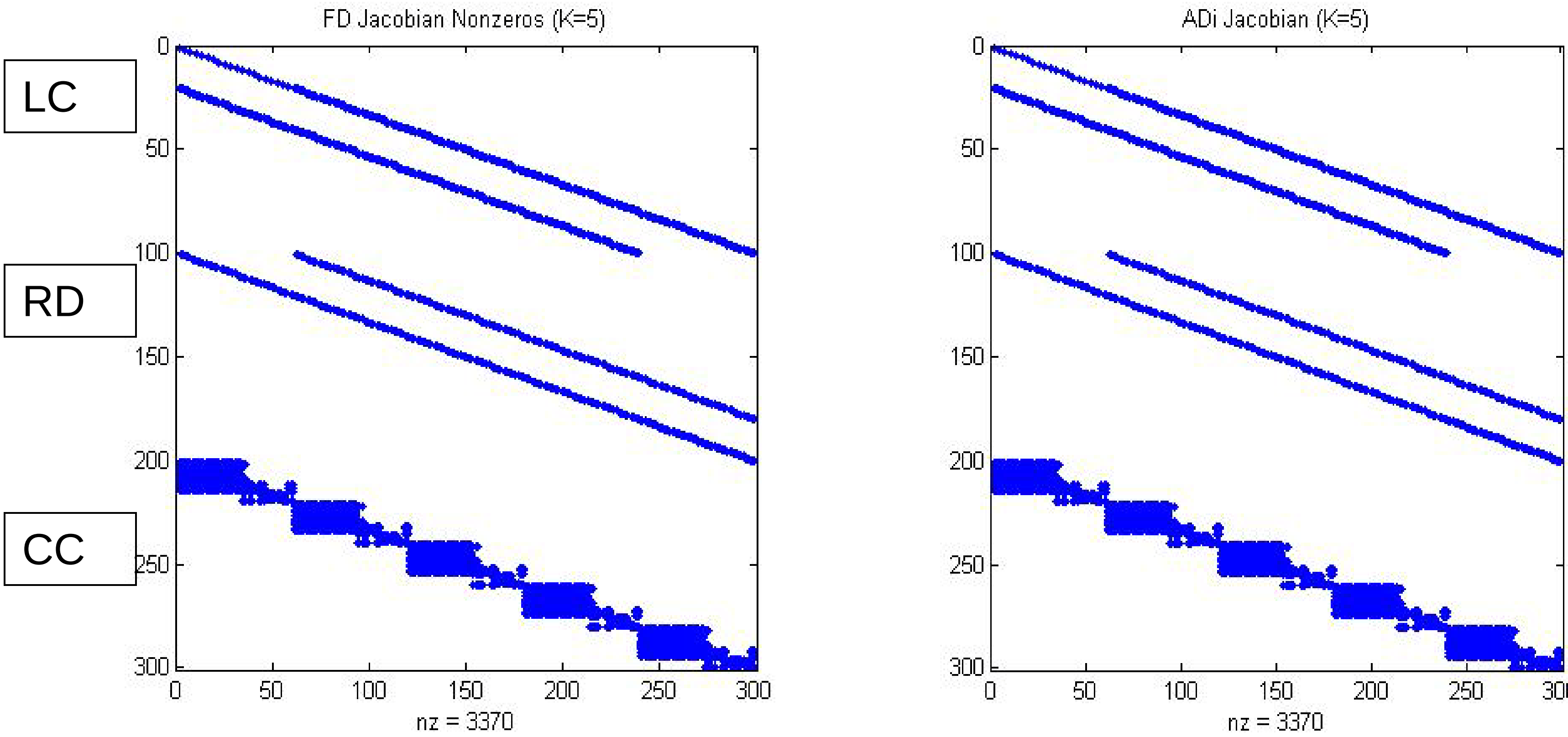
$$\tilde{y}'_{k,j}(m_k) = A_{k,j}\frac{\left(-\frac{h_k}{2}\right)}{\left(-\frac{h_k}{2}\right)\cdot(-h_k)} + B_{k,j}\frac{\left(\frac{h_k}{2}\right)+\left(-\frac{h_k}{2}\right)}{\left(\frac{h_k}{2}\right)\cdot\left(-\frac{h_k}{2}\right)} + C_{k,j}\frac{\left(\frac{h_k}{2}\right)}{(h_k)\cdot\left(\frac{h_k}{2}\right)}$$

$$\tilde{y}'_{k,j}(m_k) = A_{k,j}\frac{1}{(-h_k)} + C_{k,j}\frac{1}{(h_k)} = \frac{-A_{k,j}+C_{k,j}}{h_k} = f\left(m_k, \tilde{y}(m_k)\right)$$

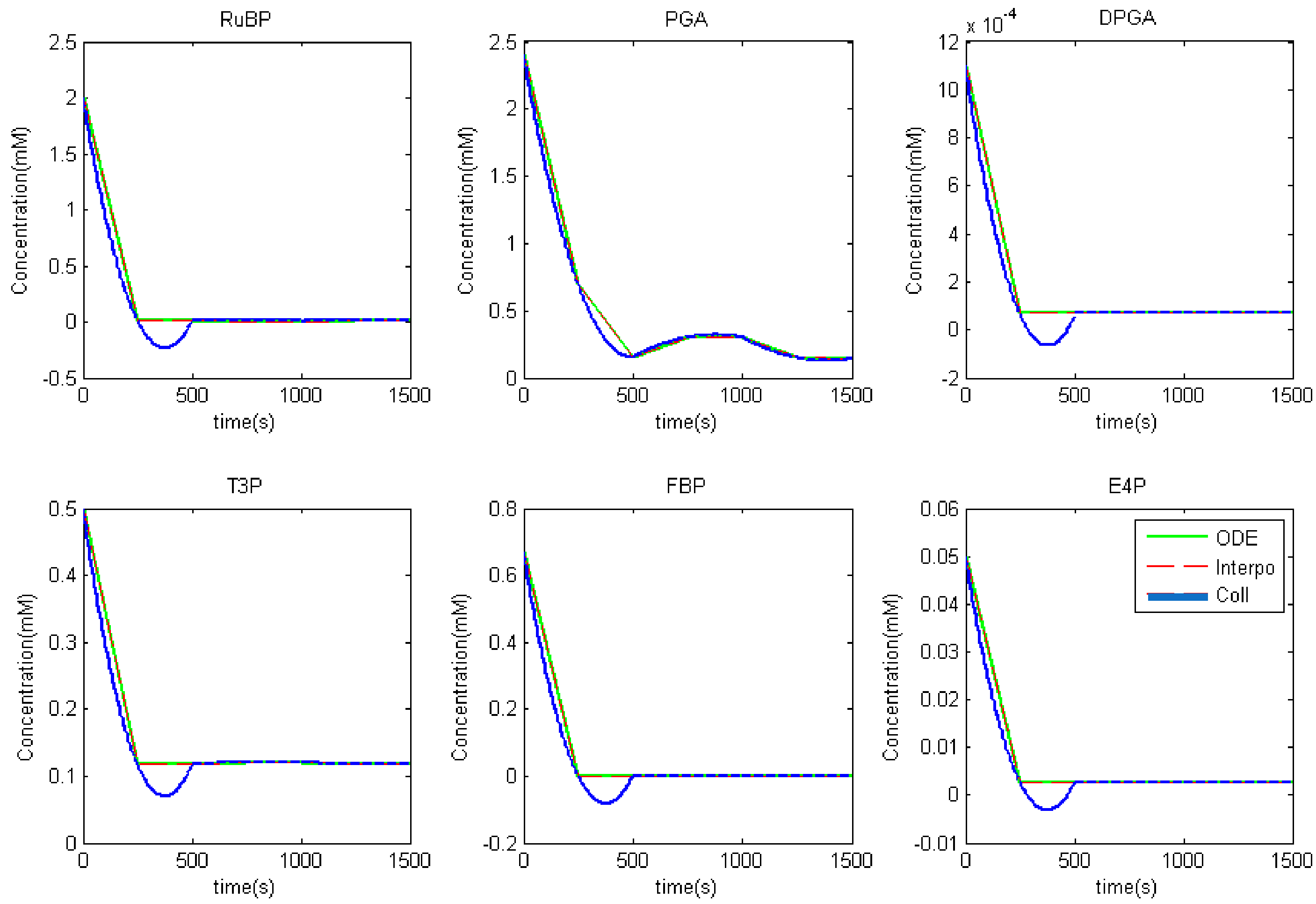
Automatic Differentiation (AD)

- Systematic application of the chain rule.
- "NUMERICALLY EXACT evaluation of derivatives"
  Shampine, Kierzenka, Forth (2005)
- Accurate to within roundoff.
- No Discretization or Cancellation errors Unlike FD
  - ADMAT/ADMIT: Verma, Coleman
  - MAD AD: Forward Mode AD [Forth, 2006].
    - Quicker w/AD than w/o AD for vectorized cases*.
  - ADIMAT: Vehreschild [18]
    - Forward AD: Directional derivatives and its Value calculated
    - More efficient than other AD [Forth, 2006]

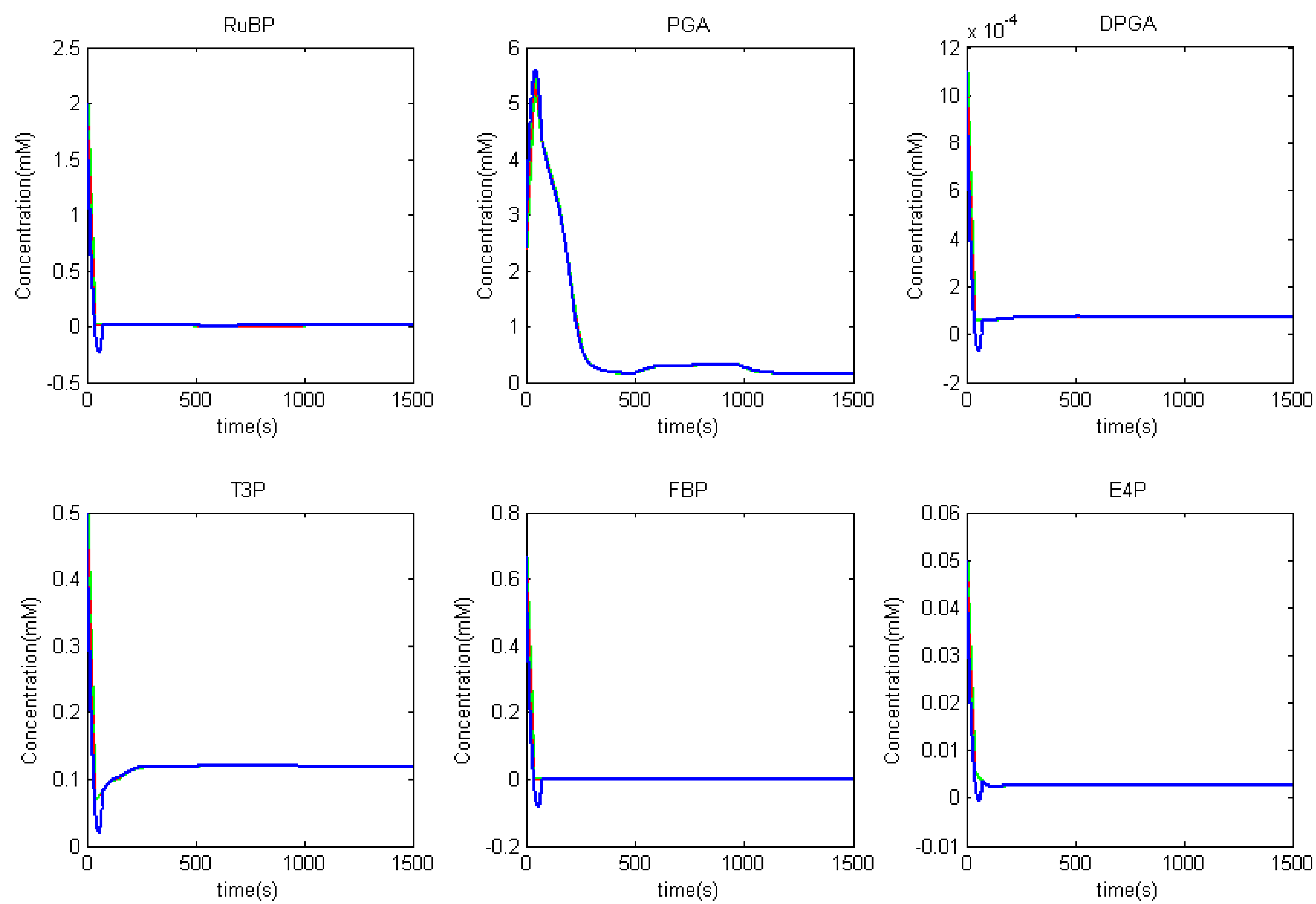
Jacobian Structure Comparison for Solving Non-linear equations. FD vs. AD



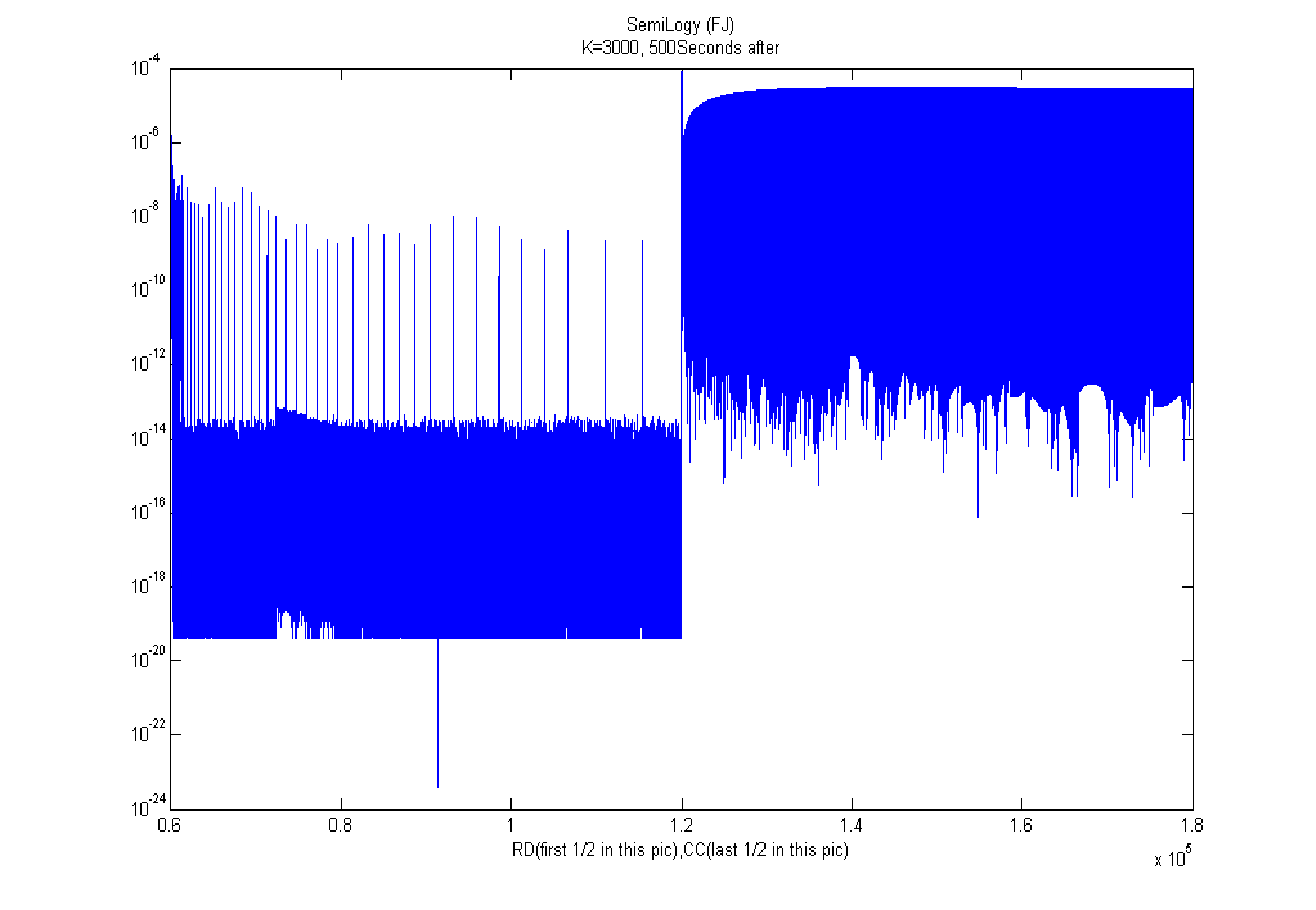
Data VS Collocation Solutions (K=3)



Data VS Collocation Solutions I (K=20)



Solving Non-Linear K=3000



Error Control and Adaptive Refinement

1.Order p: Truncation error $\tau_{i+1}(h) = O(h^p)$

2.Order $p+1$: Truncation error $\hat{\tau}_{i+1}(h) = O(h^{p+1})$

Assume $\tilde{y}_i \approx y(t_i) \approx \hat{y}_i$.

$$\tau_{i+1}(h_i) = \frac{y(t_{i+1}) - y(t_i)}{h_i} - \phi(t_i, y(t_i), h) = \frac{y(t_{i+1}) - \tilde{y}_i}{h_i} - \frac{h_i \phi(t_i, y(t_i), h)}{h_i}$$

$$= \frac{y(t_{i+1}) - [\tilde{y}_i + h_i \phi(t_i, y(t_i), h)]}{h_i} = \frac{y(t_{i+1}) - \tilde{y}_{i+1}}{h_i}$$

Similarly, $\hat{\tau}_{i+1}(h_i) = \dfrac{y(t_{i+1}) - \hat{y}_{i+1}}{h_i}$, but $\tau_{i+1}(h_i) = \dfrac{y(t_{i+1}) - \tilde{y}_{i+1}}{h_i} = \dfrac{y(t_{i+1}) - \hat{y}_{i+1}}{h_i}$

$$= \frac{y(t_{i+1}) - \hat{y}_{i+1} + (\hat{y}_{i+1} - \tilde{y}_{i+1})}{h_i} = \hat{\tau}_{i+1}(h) + \frac{\hat{y}_{i+1} - \tilde{y}_{i+1}}{h_i} \approx \frac{[\hat{y}_{i+1} - \tilde{y}_{i+1}]}{h_i} = h_i^{-1}[\hat{y}_{i+1} - \tilde{y}_{i+1}]$$

3. Given Step Size h_i, accept the step if $|\hat{y}_{i+1} - \tilde{y}_{i+1}| \le h_i Tol$

[1,4]

Nonlinear Least-Squares Problem

1. Collocation Approximation Solution.
2. Data
3. Minimize the Residual (Discrepancy btn 1 & 2).

↓

- ❑ The Gauss-Newton Method.
- ❑ The Levenberg-Marquardt Method

→ Optimized Parameters.

Why Difficult? Issues?

- Jacobians : Solving nonlinear equation & optimization.
  - Exploit Structures of Jacobians:
    - Banded FD, Block FD.
    - Coll Cond : ADiMat.
  - Vectorization is used for Jacobian build for speed.
  - Ill-Conditioned Jacobian.
    - Used different basis functions
- Stiffness of ODE:
- Sloppiness of Parameters.
- A Few Unstable Metabolites → Error ↑
  - Adaptive Step Size Refinement.

What To Do

- More efficient & robust algorithm
  - as a method of parameter estimation for broader practical applications.
  - Acceleration techniques to improve the accuracy:
- Parameter Correlation
  - Assumed zero parameter correlation now
- Combining deterministic and stochastic parameter estimation for the same model

References

[1] Ascher,U.,Mattheij,R.,and Russell, R. **Numerical Solution of Boundary Value Problems for Ordinary Differential Equations**, SIAM, Philadelphia, PA,1995.
[2] Ashyraliyev, M., Jaeger, J., Blom, J. **Parameter estimation and determinability analysis applied to Drosophila gap circuits**. 2008, BMC Systems Biology 2008 2:83
[3] Ashyraliyev, M.,Fomekong-Nanfack, Y., Kaandorp, J,, Blom, J. **Systems biology: parameter estimation for biochemical models**. 2009, FEBS Journal 276 ,2009. 886-902.
[4] Burden, R. Faires, J. **Numerical Analysis**, 7th Edition, Brooks Cole, 1997.
[5] Campbell,N.,Reece,J. **Biology**, Benjamin Cummings, 2002,pp176-196.
[6] Gautschi, w. **Numerical analysis: an introduction,** Birkhäuser, Boston, 1997.
[7] Gunkunst, R., Waterfall,J.,Casey,F. Brown, K.,Myers,C.m,Sethna,J., **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. PLOS computational Biology. 2007
[8] Heath, M. **Scientific Computing** Second Edition, McGraw Hill. New York, 2002.
[9] Iserles, A. **Numerical Analysis of Differential Equations**. Cambridge. 1996
[10] Landis, F. **Runge-Kutta and Collocation Methods**. On-line notes.
[11] Ljung, L., **Perspectives on System Identification**,2008
[12] Mendes, P,Kell, D. **Non-linear Optimization of Biochemical Pathways: applications to metabolic engineering and parameter estimation** Bioinformatics. 1998
[13] Nocedal, J., Wright, S., **Numerical Optimization, Springer**, 1999, pp
[14] Shampine,L.,Kierzenka,J., and Forth,S., **Using AD to Solve BVPs in MATLAB**. ACM Transactions on Mathematical Software.V31,No.1, 79-94. 2005
[15] Shampine,L.,Kierzenka,J., and Reichelt,M., **Solving Boundary Value Problems for Ordinary Differential Equations in MATLAB with bvp4c**. ACM Transactions on Mathematical Software. V31,No.1,79-94. 2000
[16] Stoer, J., Bulirsch, R., **Introduction to Numerical Analysis**, 2nd Edition, Springer. 1993.
[17] Tarantola, A., **Inverse Problem Theory, Methods for Data Fitting and Model Parameter Estimation.** Elsevier, 1987
[18] Vehreschild,A**., ADiMAT**.
[19] Zhu, X, Govindjee, Baker, N., de Sturler, E., Ort, D., Long, S., **Chlorophyll a fluorescence induction kinetics in leaves predicted from a model describing each discrete step of excitation energy and electron transfer associated with Photosystem II,** Planta 223(1), pages 114-133, 2005
[20] Zhu,X., de Sturler,E., Long,S. **Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm.** Plant Physiology,145(2007):513-526.
[21] Zhu,X., de Sturler,E. ,Fettig,J**. Impacts of Different Metabolite Measurement Protocols on Estimating Parameters in Complex Kinetic Metabolism models**
[22] Zhu,X., de Sturler,E., Alba,R**. A Simple Model of the Calvin Cycle has Only One Physiologically Feasible Steady State Under the Same External Conditions.** Nonlinear Analysis: Real World Application, 2008.
[23] Burns, J. Class note for **Sensitivity Analysis**, Virginia Tech, 2008.
[24] Stortelder, W., **Parameter Estimation in Nonlinear Dynamical Systems.** PhD thesis. University of Amsterdam. Mathematics and Computer Science.1998.
[25] Science Daily "**Scientists Successfully Simulate Photosynthesis And Design A Better Leaf**." 11 November 2007. <http://www.sciencedaily.com-releases/2007/11/071109123201.htm>.
[26] Sharkey T. **Photorespiration**, John Wiley & Sons, 2001.
[27] Image 1 http://www.cam.k12.il.us/hs/teachers/caplingerd/photosynthesis.gif
[28] Image 2 http://www.cosmosmagazine.com/files/imagecache/news/files/20070817_biofuel.jpg.
[29] Image 3 http://lovingthebigisland.files.wordpress.com/2009/04/eferal-sugar-cane.jpg
[30] Image 4 http://www.blogcdn.com/green.autoblog.com/media/2009/05/mobil-e85-sign-a.png
[31] E85 http://www.e85refueling.com/

Thank You So Much!