A review of sparsity vs stability in LU updates

Michael Saunders
Dept of Management Science and Engineering (MS&E)
Systems Optimization Laboratory (SOL)
Institute for Computational Mathematics and Engineering (ICME)
Stanford University, California

Session: 40 Years of Forrest and Tomlin
INFORMS 2012 Annual Meeting
Phoenix AZ, Oct 14–18, 2012
Abstract

The Forrest-Tomlin update has stood the test of time within many generations of commercial mathematical programming systems. Its ease of implementation leads to high efficiency and evidently acceptable reliability. We review its relation to Reid’s version of the Bartels-Golub update as implemented in LA05, LA15, and LUSOL. In particular, we examine the extent to which FT implementations must “live dangerously” in order to achieve the desired efficiency.
Outline

1. The Bartels-Golub update
2. The Forrest-Tomlin update
3. Sparse Bartels-Golub
4. Two tolerances
5. A numerical experiment
6. Numerical results
7. Conclusion
The Bartels-Golub update
Dense Bartels-Golub
Dense Bartels-Golub
Dense Bartels-Golub

Forget the Hessenberg
The Forrest-Tomlin update
Forrest-Tomlin (conceptually)

Few nonzeros to eliminate

$U$ stored by columns
Sparse Bartels-Golub
Sparse Bartels-Golub (Reid, Saunders)
Sparse Bartels-Golub (Reid, Saunders)

LA05, LA15, LUSOL

$U$ stored by rows
The connection

FT: Eliminate X’s with diags as pivots
BG: Allow row interchanges

Use a product of \( \begin{pmatrix} 1 & \mu \\ \mu & 1 \end{pmatrix} \) and permutations

How big can \( \mu \) safely be??
A Tale of Two Tolerances

LU factor tol $\alpha$
LU update tol $\beta$

Control and/or monitor stability
**LU factor tol** $\alpha$

Threshold partial pivoting controls $\text{cond}(L)$

$$B = LU \quad L_{ii} = 1 \quad |L_{ij}| \leq \alpha$$

$\alpha = 10.0$ or $100.0 \Rightarrow$ usually sparse and stable
LU factor tol $\alpha$

Threshold partial pivoting controls $\text{cond}(L)$

$$B = LU \quad L_{ii} = 1 \quad |L_{ij}| \leq \alpha$$

$\alpha = 10.0$ or $100.0 \Rightarrow$ usually sparse and stable

Threshold rook pivoting controls $\text{cond}(L)$, $\text{cond}(U)$

$$B = LDU \quad L_{ii} = 1 \quad |L_{ij}| \leq \alpha$$

and $U_{ii} = 1 \quad |U_{ij}| \leq \alpha$

$1 < \alpha \leq 2.0 \Rightarrow$ rank-revealing (for basis repair)
Factor and Update involve triangular matrices

\[ M = \begin{pmatrix} 1 & 1 \\ 100 & 1 \end{pmatrix} \quad \text{cond}(M) = ?? \]
Factor and Update involve triangular matrices

\[ M = \begin{pmatrix} 1 & \frac{1}{100} \\ \frac{1}{100} & 1 \end{pmatrix} \quad \text{cond}(M) \approx 10^4 \]
Factor and Update involve triangular matrices

\[ M = \begin{pmatrix} 1 & \mu \\ \mu & 1 \end{pmatrix} \quad \text{cond}(M) \approx \mu^2 \]

\[ |\mu| > 1 \]
FT update involves triangular matrices

\[ R = \begin{pmatrix}
1 &  &  & \\
 & 1 &  & \\
 &  & \ddots & \\
 &  &  & 1 \\
 & r_{p+1} & \ldots & r_m & 1
\end{pmatrix} \]

\[ \text{cond}(R) \approx r_{\text{max}}^2 \]
Product-form update involves triangular matrices

\[
T = \begin{pmatrix}
1 & v_1 \\
1 & : \\
v_p & \\
: & \ddots \\
v_m & \\
\end{pmatrix}
\]

\[
\text{cond}(T) \approx \frac{v_{\text{max}}^2}{|v_p|}
\]
LU update tol $\beta$

\[ \tilde{B} = B(I + (v - e_p)e_p^T) \]

Use $\beta$ to control or monitor basis updates

Product-form  \[ \tilde{B} = BT \]  monitor cond($T$)

Bartels-Golub  \[ \tilde{L} = LM_1M_2 \ldots \]  control cond($M_j$)

Forrest-Tomlin  \[ \tilde{L} = LR \]  monitor cond($R$)
LU update tol $\beta$

Product-form update

$$\bar{B} = BT \quad T = I + (v - e_p)e_p^T \quad \text{monitor cond}(T)$$

Update if $v_{\text{max}}^2 / |v_p| \leq \beta$ else refactor

How big should $\beta$ be?? $10^6$?
**LU update tol** $\beta$

**Product-form update**

$$\bar{B} = BT \quad T = I + (\nu - e_p)e_p^T \quad \text{monitor cond}(T)$$

Update if $\nu_{\text{max}}^2/|\nu_p| \leq \beta$ else refactor

How big should $\beta$ be?? $10^6$?

**LU updates**

<table>
<thead>
<tr>
<th>Method</th>
<th>Equation</th>
<th>Control</th>
<th>Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG</td>
<td>$\bar{L} = LM_1 M_2 \ldots$</td>
<td>$</td>
<td>M_{ij}</td>
</tr>
<tr>
<td>FT</td>
<td>$\bar{L} = LR$</td>
<td>$</td>
<td>M_{ij}</td>
</tr>
</tbody>
</table>

For BG we can enforce $\beta = 10$ say

For FT how big should $\beta$ be?? $10^6$?
A numerical experiment
Simulate FT via BG

- Run MINOS on a numerically challenging LP (e.g. pilot87)
- Set LUSOL’s LU update tol $\beta = 10, 10^2, 10^3, 10^4, 10^5, 10^6, \ldots$
- Count number of times something broke
Simulate FT via BG

- Run MINOS on a numerically challenging LP (e.g. pilot87)
- Set LUSOL’s LU update tol $\beta = 10, 10^2, 10^3, 10^4, 10^5, 10^6, \ldots$
- Count number of times something broke

Normally

- Factorization freq = 100
- LU factor tol $\alpha = 100.0$
- LU update tol $\beta = 10.0$
Simulate FT via BG

- Run MINOS on a numerically challenging LP (e.g. pilot87)
- Set LUSOL’s **LU update tol** $\beta = 10, 10^2, 10^3, 10^4, 10^5, 10^6, \ldots$
- Count number of times something **broke**

Normally

- Factorization freq $= 100$
- LU factor tol $\alpha = 100.0$
- LU update tol $\beta = 10.0$

- How would MINOS measure **broke**?
Simulate FT via BG

- Run MINOS on a numerically challenging LP (e.g. pilot87)
- Set LUSOL’s LU update tol $\beta = 10, 10^2, 10^3, 10^4, 10^5, 10^6, \ldots$
- Count number of times something broke

Normally
- Factorization freq $= 100$
- LU factor tol $\alpha = 100.0$
- LU update tol $\beta = 10.0$

- How would MINOS measure broke?
- Row check failed? Every 60 itns, see if $\|Ax - b\|$ is too big
  - Never happened
Simulate FT via BG

- Run MINOS on a numerically challenging LP (e.g. pilot87)
- Set LUSOL’s LU update tol $\beta = 10, 10^2, 10^3, 10^4, 10^5, 10^6, \ldots$
- Count number of times something broke

Normally

- Factorization freq = 100
- LU factor tol $\alpha = 100.0$
- LU update tol $\beta = 10.0$

- How would MINOS measure broke?
- Row check failed? Every 60 itns, see if $\|Ax - b\|$ is too big
  Never happened
- LU became singular? Every update, check if $U$ has small diag
  Increasingly often
Tests for accepting LU update

\[ \tilde{L} = \begin{pmatrix} 1 & \mathbf{r}^T \\ \mathbf{r} & 1 \end{pmatrix} \quad \tilde{U} = \begin{pmatrix} U_1 & \mathbf{u} \\ \delta \end{pmatrix} \]

**Tomlin 1975**

\[ \frac{r_{\text{max}} u_{\text{max}}}{|\delta|} \leq 16^7 \quad (3 \cdot 10^8) \]

**Our experiment**

\[ r_{\text{max}} \leq \beta \quad \text{and} \quad \frac{\|\mathbf{u}\|_1}{|\delta|} \leq \epsilon^{-2/3} \quad (3 \cdot 10^{10}) \]
Tests for accepting LU update

\[ \bar{L} = L \begin{pmatrix} I \\ r^T \\ 1 \end{pmatrix} \quad \bar{U} = \begin{pmatrix} U_1 & u \\ \delta \end{pmatrix} \]

Tomlin 1975

\[ \frac{r_{\text{max}} u_{\text{max}}}{|\delta|} \leq 3 \cdot 10^8 \]

Our experiment

\[ r_{\text{max}} \leq \beta \quad \text{and} \quad \frac{||u||_1}{|\delta|} \leq \epsilon^{-2/3} \quad (3 \cdot 10^{10}) \]

Roughly equivalent to

\[ \frac{r_{\text{max}} u_{\text{max}}}{|\delta|} \leq \beta 10^{10} = 10^{11}, 10^{12}, \ldots \]
Numerical results
MINOS simulation of FT updates

LPnetlib problem pilot $\epsilon = 10^{-16}$ Featol = Opttol = $10^{-6}$
LU update tol $\beta = 10, 10^2, 10^3, \ldots$

<table>
<thead>
<tr>
<th>beta</th>
<th>itns</th>
<th>time</th>
<th>LUfac</th>
<th>warnings</th>
<th>average updates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1e+1$</td>
<td>16865</td>
<td>5.06</td>
<td>167</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>$1e+2$</td>
<td>16865</td>
<td>5.02</td>
<td>170</td>
<td>19</td>
<td>99</td>
</tr>
<tr>
<td>$1e+3$</td>
<td>16865</td>
<td>5.24</td>
<td>220</td>
<td>176</td>
<td>77</td>
</tr>
<tr>
<td>$1e+4$</td>
<td>16865</td>
<td>5.73</td>
<td>305</td>
<td>287</td>
<td>55</td>
</tr>
<tr>
<td>$1e+5$</td>
<td>18837</td>
<td>7.05</td>
<td>445</td>
<td>442</td>
<td>42</td>
</tr>
<tr>
<td>$1e+6$</td>
<td>20079</td>
<td>8.28</td>
<td>576</td>
<td>573</td>
<td>35</td>
</tr>
<tr>
<td>$1e+7$</td>
<td>16704</td>
<td>7.21</td>
<td>514</td>
<td>510</td>
<td>32</td>
</tr>
<tr>
<td>$1e+8$</td>
<td>21115</td>
<td>9.31</td>
<td>678</td>
<td>674</td>
<td>31</td>
</tr>
</tbody>
</table>
MINOS simulation of FT updates

pilots: LU update tol = 1e+8  Average 31 updates

Michael Saunders

Sparsity vs stability in LU updates
MINOS simulation of FT updates

LPnetlib problem **pilot87** \( \epsilon = 10^{-16} \) \( \text{Featol} = \text{Opttol} = 10^{-6} \)

LU update tol \( \beta = 10, 10^2, 10^3, \ldots \)

<table>
<thead>
<tr>
<th>beta</th>
<th>itns</th>
<th>time</th>
<th>LUfac</th>
<th>warnings</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e+1</td>
<td>20055</td>
<td>14.64</td>
<td>197</td>
<td>0</td>
<td>101</td>
</tr>
<tr>
<td>1e+2</td>
<td>19159</td>
<td>13.99</td>
<td>194</td>
<td>30</td>
<td>99</td>
</tr>
<tr>
<td>1e+3</td>
<td>20001</td>
<td>15.34</td>
<td>259</td>
<td>202</td>
<td>77</td>
</tr>
<tr>
<td>1e+4</td>
<td>20001</td>
<td>17.66</td>
<td>380</td>
<td>360</td>
<td>53</td>
</tr>
<tr>
<td>1e+5</td>
<td>21635</td>
<td>22.19</td>
<td>526</td>
<td>511</td>
<td>41</td>
</tr>
<tr>
<td>1e+6</td>
<td>21837</td>
<td>24.41</td>
<td>624</td>
<td>617</td>
<td>35</td>
</tr>
<tr>
<td>1e+7</td>
<td>23158</td>
<td>27.72</td>
<td>772</td>
<td>770</td>
<td>30</td>
</tr>
<tr>
<td>1e+8</td>
<td>36862</td>
<td>48.91</td>
<td>1385</td>
<td>1379</td>
<td>27</td>
</tr>
</tbody>
</table>
MINOS simulation of FT updates

LPnetlib problem *mod2pre* \(\epsilon = 10^{-16}\)  \(\text{Featol} = \text{Opttol} = 10^{-6}\)

LU update tol \(\beta = 10, 10^2, 10^3, \ldots\)

<table>
<thead>
<tr>
<th>(\beta)</th>
<th>itns</th>
<th>time</th>
<th>LUfac</th>
<th>warnings</th>
<th>singular</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e+1</td>
<td>500000</td>
<td>705.0</td>
<td>4844</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1e+2</td>
<td>500000</td>
<td>704.9</td>
<td>4859</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>1e+3</td>
<td>500000</td>
<td>698.5</td>
<td>5312</td>
<td>1860</td>
<td></td>
</tr>
<tr>
<td>1e+4</td>
<td>11057*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e+5</td>
<td>17256*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e+6</td>
<td>500000</td>
<td>793.4</td>
<td>8955</td>
<td>7941</td>
<td>859</td>
</tr>
<tr>
<td>1e+7</td>
<td>99774*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e+8</td>
<td>500000</td>
<td>785.6</td>
<td>7528</td>
<td>5661</td>
<td>3000</td>
</tr>
</tbody>
</table>

*Refactorization gave singular LU
Rook pivoting invoked, \(\beta\) reduced to 10.0
Quad SQOPT simulation of FT updates

LPnetlib problem pilot $\epsilon = 10^{-34}$  Featol = Opttol = $10^{-15}$
LU update tol $\beta = 10^4, 10^6, 10^8, \ldots$

<table>
<thead>
<tr>
<th>beta</th>
<th>itns</th>
<th>time</th>
<th>LUfacs</th>
<th>singular</th>
<th>average</th>
<th>dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e+4</td>
<td>17251</td>
<td>216.0</td>
<td>173</td>
<td>2</td>
<td>100</td>
<td>E-25</td>
</tr>
<tr>
<td>1e+6</td>
<td>17251</td>
<td>215.4</td>
<td>187</td>
<td>70</td>
<td>92</td>
<td>E-27</td>
</tr>
<tr>
<td>1e+8</td>
<td>17251</td>
<td>215.1</td>
<td>213</td>
<td>144</td>
<td>81</td>
<td>E-21</td>
</tr>
<tr>
<td>1e+10</td>
<td>17251</td>
<td>216.0</td>
<td>252</td>
<td>205</td>
<td>68</td>
<td>E-24</td>
</tr>
<tr>
<td>1e+12</td>
<td>17251</td>
<td>220.3</td>
<td>264</td>
<td>219</td>
<td>65</td>
<td>E-27</td>
</tr>
<tr>
<td>1e+14</td>
<td>17251</td>
<td>219.8</td>
<td>264</td>
<td>206</td>
<td>65</td>
<td>E-29</td>
</tr>
<tr>
<td>1e+16</td>
<td>17251</td>
<td>217.6</td>
<td>229</td>
<td>150</td>
<td>75</td>
<td>E-12</td>
</tr>
<tr>
<td>1e+18</td>
<td>17251</td>
<td>216.9</td>
<td>212</td>
<td>111</td>
<td>81</td>
<td>E-19</td>
</tr>
<tr>
<td>1e+20</td>
<td>11013*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e+22</td>
<td>7864*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e+24</td>
<td>4098*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Refac singular, beta reduced to 50.0
Conclusion
Conclusion

Test for accepting an FT update

| WHIZARD       | \( \frac{r_{\text{max}} u_{\text{max}}}{|\delta|} \) \leq 10^8 |
|---------------|----------------------------------------------------------|
| CLP, CPLEX, Gurobi, Xpress, Mozek, ... | \( \frac{r_{\text{max}} u_{\text{max}}}{|\delta|} \) \leq ? or ? |
| Probably safer | \( r_{\text{max}} \leq \beta \) and \( \frac{u_{\text{max}}}{|\delta|} \leq 10^{10} \) |

Average updates per LU

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>Updates</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^4 )</td>
<td>50</td>
</tr>
<tr>
<td>( 10^5 )</td>
<td>40</td>
</tr>
<tr>
<td>( 10^6 )</td>
<td>35</td>
</tr>
<tr>
<td>( 10^7 )</td>
<td>30</td>
</tr>
</tbody>
</table>
References (in chronological order)

  Eliminated subdiag of dense Hessenberg $H = \bar{U}P$

Michael Saunders

Sparsity vs stability in LU updates
References (in chronological order)


References (in chronological order)

  Eliminated subdiag of dense Hessenberg $H = \bar{U}P$

  Eliminated $p$th row of sparse $H$ (not subdiag)

  Test for an unstable update
References (in chronological order)

  Eliminated subdiag of dense Hessenberg $H = \bar{U}P$

  Eliminated $p$th row of sparse $H$ (not subdiag)

  Test for an unstable update

  Picture of symmetrically permuted $P^TU\bar{U}P$ with row spike to be eliminated

*U* stored by rows to facilitate BG update
References (contd)

  *U* stored by rows to facilitate BG update

  Rectangular LU, more general updates, cond(*L*) controlled throughout
References (contd)

  
  $U$ stored by rows to facilitate BG update

  
  Rectangular LU, more general updates, $\text{cond}(L)$ controlled throughout

  
  LU with Threshold Rook Pivoting, same general updates
Forrest-Tomlin