Nonlinear Least-Squares Problems with the Gauss-Newton and Levenberg-Marquardt Methods

Alfonso Croeze[1] Lindsey Pittman[2] Winnie Reynolds[1]

[1]Department of Mathematics
Louisiana State University
Baton Rouge, LA

[2]Department of Mathematics
University of Mississippi
Oxford, MS

July 6, 2012

Optimization

- The process of finding the minimum or maximum value of an objective function (e.g. maximizing profit, minimizing cost).
- Constrained or unconstrained.
- Useful in nonlinear least-squares problems.

Terminology I

The *gradient* ∇f of a multivariable function is a vector consisting of the function's partial derivatives:

$$\nabla f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)$$

The *Hessian matrix* $H(f)$ of a function $f(x)$ is the square matrix of second-order partial derivatives of $f(x)$:

$$H(f(x_1, x_2)) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1^2} & \dfrac{\partial f}{\partial x_1 \partial x_2} \\ \dfrac{\partial f}{\partial x_1 \partial x_2} & \dfrac{\partial f}{\partial x_2^2} \end{pmatrix}$$

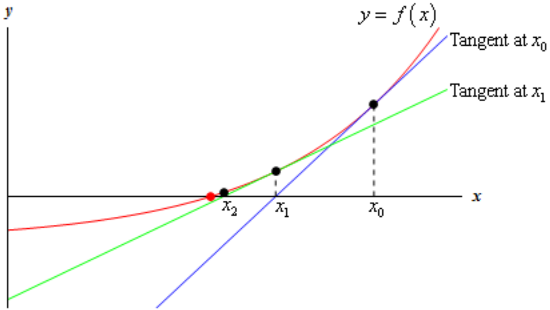
Terminology II

The *transpose* $A^\top$ of a matrix A is the matrix created by reflecting A over its main diagonal:

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}^\top = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Matrix A is *positive-definite* if, for all real non-zero vectors z, $z^\top Az > 0$.

Newton's Method



$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \left(= x_n - \frac{f'(x_n)}{f''(x_n)} \right)$$

Nonlinear Least-Squares I

- A form of regression where the objective function is the sum of squares of nonlinear functions:

$$f(x) = \frac{1}{2}\sum_{j=1}^{m}(r_j(x))^2 = \frac{1}{2}||r(x)||_2^2$$

- The j-th component of the m-vector $r(x)$ is the residual $r_j(x) = \phi(x; t_j) - y_j$:

$$r(x) = (r_1(x), r_2(x), ..., r_m(x))^T$$

Nonlinear Least-Squares II

The *Jacobian* $J(x)$ is a matrix of all $\nabla r_j(x)$:

$$J(x) = \left[\frac{\partial r_j}{\partial x_i}\right]_{j=1,\ldots,m;i=1,\ldots,n} = \begin{bmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix}$$

Nonlinear Least-Squares III

The gradient and Hessian of $f(x)$ can be expressed in terms of the Jacobian:

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x) \nabla r_j(x) = J(x)^T r(x)$$

$$\begin{aligned}\nabla^2 f(x) &= \sum_{j=1}^{m} \nabla r_j(x) \nabla r_j(x)^T + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x) \\ &= J(x)^T J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x)\end{aligned}$$

The Gauss-Newton Method I

- Generalizes Newton's method for multiple dimensions
- Uses a line search: $x_{k+1} = x_k + \alpha_k p_k$
- The values being altered are the variables of the model $\phi(x; t_j)$

The Gauss-Newton Method II

- Replace $f'(x)$ with the gradient ∇f
- Replace $f''(x)$ with the Hessian $\nabla^2 f$
- Use the approximation $\nabla^2 f_k \approx J_k^T J_k$

$$J_k^T J_k p_k^{GN} = -J_k^T r_k$$

- J_k must have full rank
- Requires accurate initial guess
- Fast convergence close to solution

The Gauss-Newton Method III

```
x = {2.50, 0.25}
f[x1_, x2_] := x1*e^(x2*t) - y
g[x1_, x2_] := D[f[x1, x2], {{x1, x2}}]
t = {{1}, {2}, {4}, {5}, {8}}
y = {{3.2939}, {4.2699}, {7.1749}, {9.3008}, {20.259}}

f = f[2.50, 0.25]
Jf = Flatten[Transpose[g[x1, x2] /. {x1 → 2.50, x2 → 0.25}], 1]
JR = Transpose[Transpose[f].Jf]
Hf = Transpose[Jf].Jf
p = LinearSolve[Hf , -JR]
a0 = x + p
```

The Gauss-Newton Method IV

$$x0 = \begin{pmatrix} 2.5` \\ 0.25` \end{pmatrix}$$

$$x1 = \begin{pmatrix} 2.538107 \\ 0.260178 \end{pmatrix}$$

$$x2 = \begin{pmatrix} 2.541052 \\ 0.259503 \end{pmatrix}$$

$$x3 = \begin{pmatrix} 2.541071 \\ 0.259502 \end{pmatrix}$$

$$x4 = \begin{pmatrix} 2.541070 \\ 0.259502 \end{pmatrix}$$

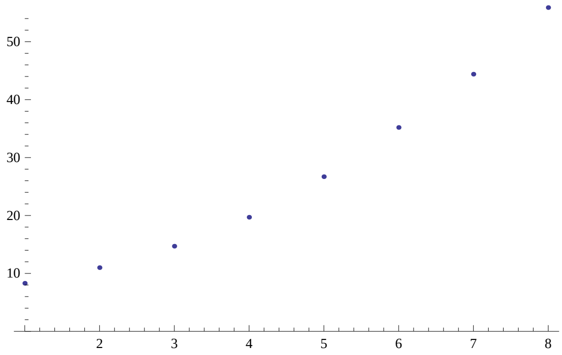
$$x5 = \begin{pmatrix} 2.541069 \\ 0.259502 \end{pmatrix}$$

GN Example: Exponential Data I

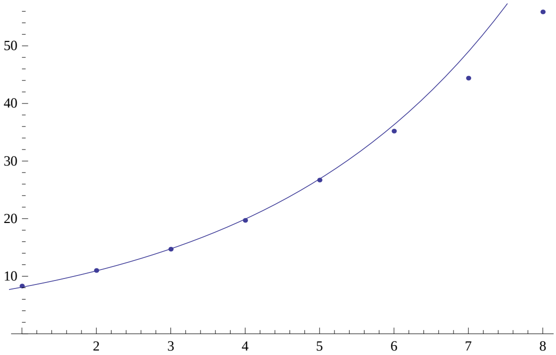
United States population (in millions) and the corresponding year:

Year	Population
1815	8.3
1825	11.0
1835	14.7
1845	19.7
1855	26.7
1865	35.2
1875	44.4
1885	55.9

GN Example: Exponential Data II



GN Example: Exponential Data III



$$\phi(x;t) = x_1 e^{x_2 t}; x_1 = 6, x_2 = .3$$

GN Example: Exponential Data IV

$$r(x) = \begin{pmatrix} 6e^{.3(1)} - 8.3 \\ 6e^{.3(2)} - 11 \\ 6e^{.3(3)} - 14.7 \\ 6e^{.3(4)} - 19.7 \\ 6e^{.3(5)} - 26.7 \\ 6e^{.3(6)} - 35.2 \\ 6e^{.3(7)} - 44.4 \\ 6e^{.3(8)} - 55.9 \end{pmatrix} = \begin{pmatrix} -0.200847 \\ -0.0672872 \\ 0.0576187 \\ 0.220702 \\ 0.190134 \\ 1.09788 \\ 4.59702 \\ 10.2391 \end{pmatrix}$$

$$||r||^2 = 127.309$$

GN Example: Exponential Data V

$$\phi(x;t) = x_1 e^{x_2 t}$$

$$J(x) = \begin{pmatrix} e^{x_2} & e^{x_2}x_1 \\ e^{2x_2} & 2e^{2x_2}x_1 \\ e^{3x_2} & 3e^{3x_2}x_1 \\ e^{4x_2} & 4e^{4x_2}x_1 \\ e^{5x_2} & 5e^{5x_2}x_1 \\ e^{6x_2} & 6e^{6x_2}x_1 \\ e^{7x_2} & 7e^{7x_2}x_1 \\ e^{8x_2} & 8e^{8x_2}x_1 \end{pmatrix} = \begin{pmatrix} 1.34986 & 8.09915 \\ 1.82212 & 21.8654 \\ 2.4596 & 44.2729 \\ 3.32012 & 79.6828 \\ 4.48169 & 134.451 \\ 6.04965 & 217.787 \\ 8.16617 & 342.979 \\ 11.0232 & 529.112 \end{pmatrix}$$

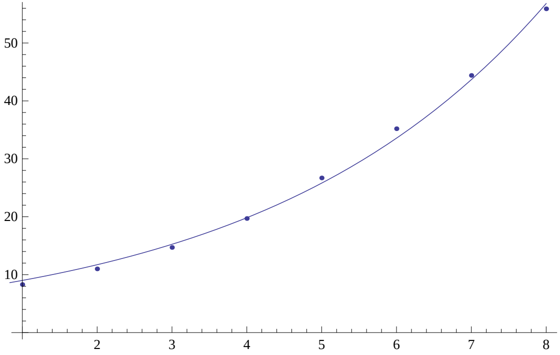
GN Example: Exponential Data VI

$$Solve[\{J1^T.J1.\{\{p1\},\{p2\}\} == -J1^T.R1\},\{p1,p2\}]$$

$$\{\{p1 \to 0.923529, p2 \to -0.0368979\}\}$$

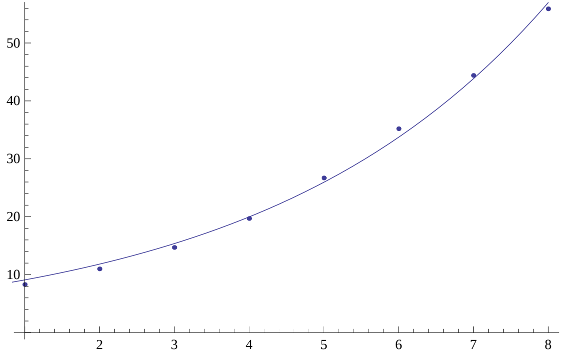
$$\begin{aligned} x_{1_1} &= 6 + 0.923529 = 6.92353 \\ x_{2_1} &= .3 - 0.0368979 = 0.263103 \end{aligned}$$

GN Example: Exponential Data VII



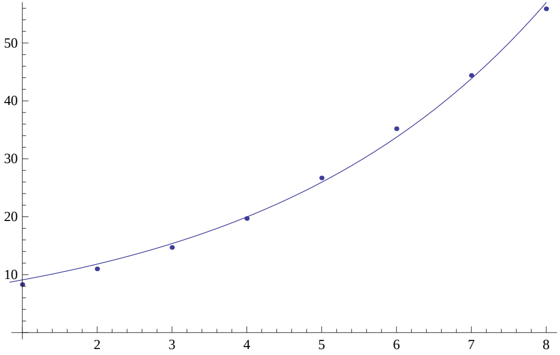
$$||r||^2 = 6.16959$$

GN Example: Exponential Data VIII



$$||r||^2 = 6.01313$$

GN Example: Exponential Data IX



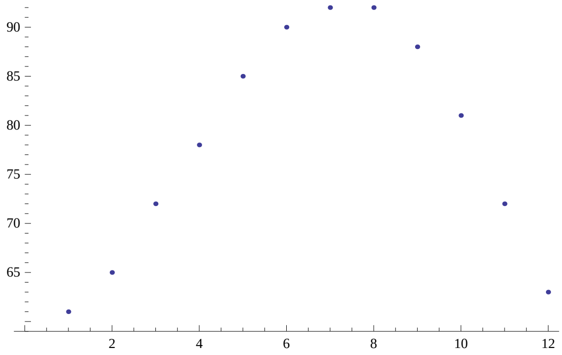
$$||r||^2 = 6.01308; x = (7.00009, 0.262078)$$

GN Example: Sinusoidal Data I

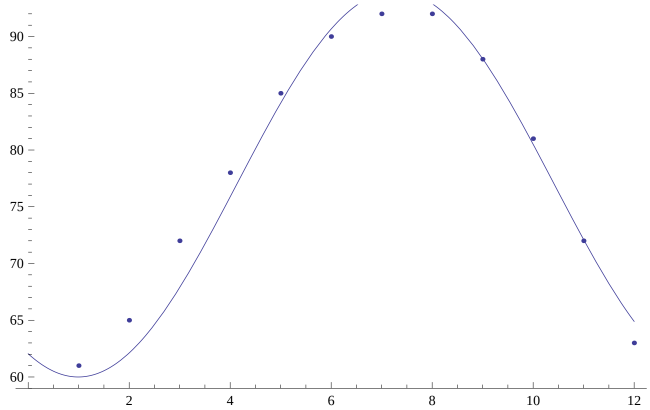
Average monthly high temperatures for Baton Rouge, LA:

Jan	61	Jul	92
Feb	65	Aug	92
Mar	72	Sep	88
Apr	78	Oct	81
May	85	Nov	72
Jun	90	Dec	63

GN Example: Sinusoidal Data II



GN Example: Sinusoidal Data III



$$\phi(x;t) = x_1 \sin(x_2 t + x_3) + x_4; x = (17, .5, 10.5, 77)$$

GN Example: Sinusoidal Data IV

$$r(x) = \begin{pmatrix} 17\sin(.5(1)+10.5)+77-61 \\ 17\sin(.5(2)+10.5)+77-65 \\ 17\sin(.5(3)+10.5)+77-72 \\ 17\sin(.5(4)+10.5)+77-78 \\ 17\sin(.5(5)+10.5)+77-85 \\ 17\sin(.5(6)+10.5)+77-90 \\ 17\sin(.5(7)+10.5)+77-92 \\ 17\sin(.5(8)+10.5)+77-92 \\ 17\sin(.5(9)+10.5)+77-88 \\ 17\sin(.5(10)+10.5)+77-81 \\ 17\sin(.5(11)+10.5)+77-72 \\ 17\sin(.5(12)+10.5)+77-63 \end{pmatrix} = \begin{pmatrix} -0.999834 \\ -2.88269 \\ -4.12174 \\ -2.12747 \\ -0.85716 \\ 0.664335 \\ 1.84033 \\ 0.893216 \\ 0.0548933 \\ -0.490053 \\ 0.105644 \\ 1.89965 \end{pmatrix}$$

$$||r||^2 = 40.0481$$

GN Example: Sinusoidal Data V

$$J(x) = \begin{pmatrix}
\sin(x_2 + x_3) & x_1 \cos(x_2 + x_3) & x_1 \cos(x_2 + x_3) & 1 \\
\sin(2x_2 + x_3) & 2x_1 \cos(2x_2 + x_3) & x_1 \cos(2x_2 + x_3) & 1 \\
\sin(3x_2 + x_3) & 3x_1 \cos(3x_2 + x_3) & x_1 \cos(3x_2 + x_3) & 1 \\
\sin(4x_2 + x_3) & 4x_1 \cos(4x_2 + x_3) & x_1 \cos(4x_2 + x_3) & 1 \\
\sin(5x_2 + x_3) & 5x_1 \cos(5x_2 + x_3) & x_1 \cos(5x_2 + x_3) & 1 \\
\sin(6x_2 + x_3) & 6x_1 \cos(6x_2 + x_3) & x_1 \cos(6x_2 + x_3) & 1 \\
\sin(7x_2 + x_3) & 7x_1 \cos(7x_2 + x_3) & x_1 \cos(7x_2 + x_3) & 1 \\
\sin(8x_2 + x_3) & 8x_1 \cos(8x_2 + x_3) & x_1 \cos(8x_2 + x_3) & 1 \\
\sin(9x_2 + x_3) & 9x_1 \cos(9x_2 + x_3) & x_1 \cos(9x_2 + x_3) & 1 \\
\sin(10x_2 + x_3) & 10x_1 \cos(10x_2 + x_3) & x_1 \cos(10x_2 + x_3) & 1 \\
\sin(11x_2 + x_3) & 11x_1 \cos(11x_2 + x_3) & x_1 \cos(11x_2 + x_3) & 1 \\
\sin(12x_2 + x_3) & 12x_1 \cos(12x_2 + x_3) & x_1 \cos(12x_2 + x_3) & 1
\end{pmatrix}$$

GN Example: Sinusoidal Data VI

$$J(x) = \begin{pmatrix} -0.99999 & 0.0752369 & 0.0752369 & 1 \\ -0.875452 & 16.4324 & 8.21618 & 1 \\ -0.536573 & 43.0366 & 14.3455 & 1 \\ -0.0663219 & 67.8503 & 16.9626 & 1 \\ 0.420167 & 77.133 & 15.4266 & 1 \\ 0.803784 & 60.6819 & 10.1137 & 1 \\ 0.990607 & 16.2717 & 2.32453 & 1 \\ 0.934895 & -48.2697 & -6.03371 & 1 \\ 0.650288 & -116.232 & -12.9147 & 1 \\ 0.206467 & -166.337 & -16.6337 & 1 \\ -0.287903 & -179.082 & -16.2802 & 1 \\ -0.711785 & -143.289 & -11.9407 & 1 \end{pmatrix}$$

GN Example: Sinusoidal Data VII

$$Solve[\{J1^T.J1.\{\{p1\},\{p2\},\{p3\},\{p4\}\} ==$$

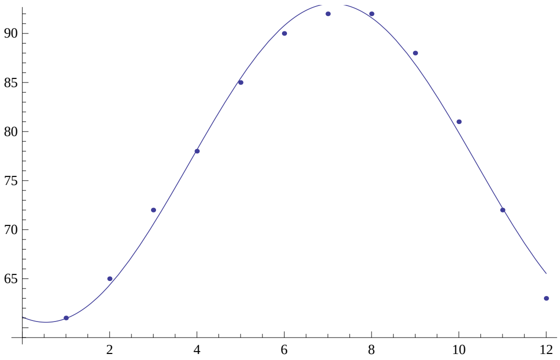
$$-J1^T.R1\},\{p1, p2, p3, p4\}]$$

$$\{\{p1 \to -0.904686, p2 \to -0.021006,$$

$$p3 \to 0.230013, p4 \to -0.17933\}\}$$

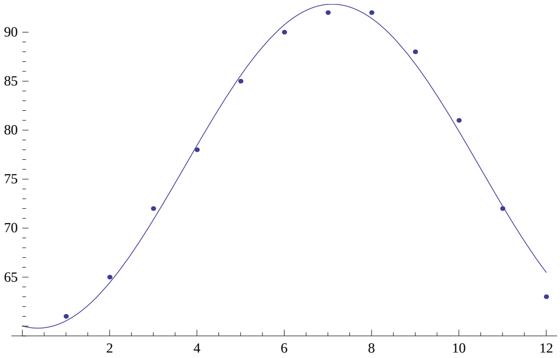
$$\begin{aligned} x_{1_1} &= 17 - 0.904686 = 16.0953 \\ x_{2_1} &= .5 - 0.021006 = 0.478994 \\ x_{3_1} &= 10.5 + 0.230013 = 10.73 \\ x_{4_1} &= 77 - 0.17933 = 76.8207 \end{aligned}$$

GN Example: Sinusoidal Data VIII



$$||r||^2 = 13.8096$$

GN Example: Sinusoidal Data IX



$$||r||^2 = 13.6556; x = (16.2411, 0.47912, 10.7335, 76.7994)$$

Gradient Descent

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k)$$

- Quickly approaches the solution from a distance.
- Convergence becomes very slow close to the solution.

The Levenberg-Marquardt Method I

- Same approximation for the Hessian matrix as GN
- Implements a trust region strategy instead of a line search technique: At each iteration we must solve

$$\min_p \frac{1}{2}||J_k p_k + r_k||^2, \qquad \text{subject to } ||p_k|| \leq \Delta_k$$

- The model function $m_k(p)$ is a restatement of the trust region equation using our approximations of $f(x)$, $\nabla f(x)$, and the Hessian of $f(x)$:

$$m_k(p) = \frac{1}{2}||r_k||^2 + p_k^T J_k^T r_k + \frac{1}{2} p_k^T J_k^T J_k p_k$$

The Levenberg-Marquardt Method II

- The value for Δ_k is chosen for each iteration depending on the error value of the corresponding p_k.
- Look at the comparison of the *actual reduction* in the numerator and the *predicted reduction* in the denominator.

$$\rho_k = \frac{d(x_k) - d(x_k + p_k)}{\phi_k(0) - \phi_k(p_k)}$$

- If ρ_k is close to 1 $\rightarrow$ expand Δ_k
- If ρ_k is positive but significantly smaller than 1 $\rightarrow$ keep Δ_k
- If ρ_k is close to zero or negative $\rightarrow$ shrink Δ_k
- Next, solve for p_k.

The Levenberg-Marquardt Method III

- If p_k^{GN} does not lie inside the trust region Δ_k, then there must be some $\lambda > 0$ such that

$$(J_k^T J_k + \lambda I)p_k^{LM} = -J_k^T r_k$$

- This new p_k^{LM} has the property $||p_k^{LM}|| = \Delta_k$.
- Typically λ_1 is chosen to be small (1). It is then altered at each iteration to find an appropriate p_k.
- We then minimize p_k
- Update variables of the model function and repeat the process, finding a new Δ_{k+1}.

LMA Example: Exponential Data I

Year	Population
1815	8.3
1825	11.0
1835	14.7
1845	19.7
1855	26.7
1865	35.2
1875	44.4
1885	55.9

$$||p_1^{GN}|| = 0.924266$$

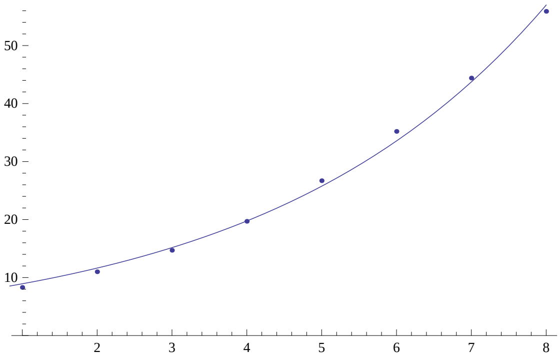
LMA Example: Exponential Data II

$$Solve[(J1^T.J1 + 1 * \{\{1,0\},\{0,1\}\}).\{\{p1\},\{p2\}\} == -J1^T.R1, \{p1, p2\}]$$

$$\{\{p1 \to 0.851068, p2 \to -0.0352124\}\}$$

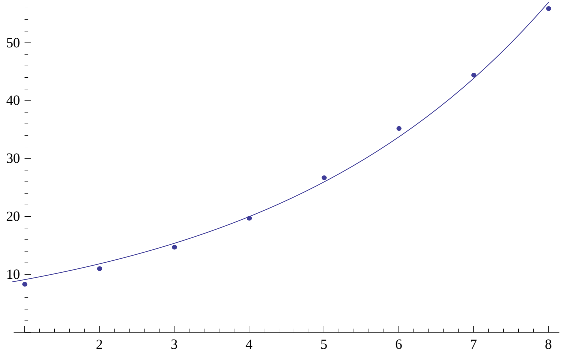
$$x_{1_1} = 6 + 0.851068 = 6.85107$$
$$x_{2_1} = .3 - 0.0352124 = 0.264788$$

LMA Example: Exponential Data III



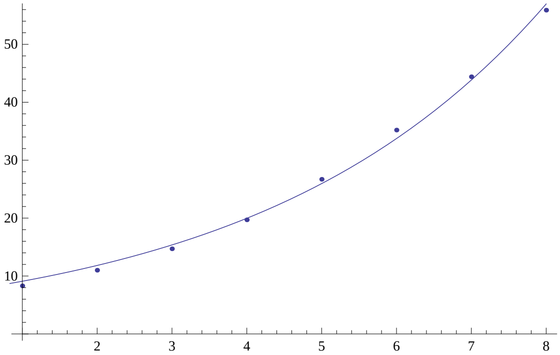
$$||r||^2 = 6.16959; ||p_1^{LM}|| = 0.851796$$

LMA Example: Exponential Data IV



$$||r||^2 = 6.01312; ||p_2^{LM}|| = 0.150782$$

LMA Example: Exponential Data V



$$||r||^2 = 6.01308; ||p_3^{LM}|| = 0.000401997$$
$$x = (7.00012, 0.262077)$$

LMA Example: Sinusoidal Data I

Jan	61	Jul	92
Feb	65	Aug	92
Mar	72	Sep	88
Apr	78	Oct	81
May	85	Nov	72
Jun	90	Dec	63

$$||p_1^{GN}|| = 0.95077$$

LMA Example: Sinusoidal Data II

$$Solve[(J1^T.J1+1*\{\{1,0,0,0\},\{0,1,0,0\},\{0,0,1,0\},\{0,0,0,1\}\}).$$

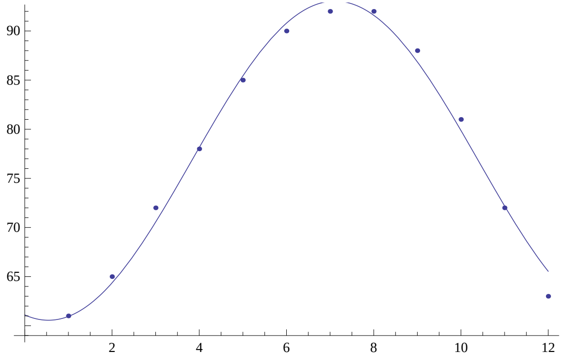
$$\{\{p1\},\{p2\},\{p3\},\{p4\}\} == -J1^T.R1,\{p1,p2,p3,p4\}]$$

$$\{\{p1 \to -0.7595, p2 \to -0.0219004,$$

$$p3 \to 0.236647, p4 \to -0.198876\}\}$$

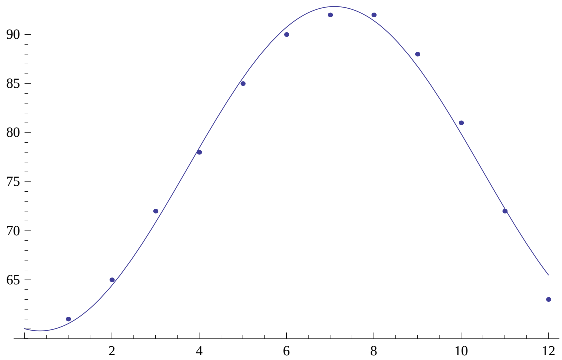
$$\begin{aligned} x_{1_1} &= 17 - 0.7595 = 16.2405 \\ x_{2_1} &= .5 - 0.0219004 = 0.4781 \\ x_{3_1} &= 10.5 + 0.236647, = 10.7366 \\ x_{4_1} &= 77 - 0.198876 = 76.8011 \end{aligned}$$

LMA Example: Sinusoidal Data III



$$||r||^2 = 13.6458; ||p_1^{LM}|| = 0.820289$$

LMA Example: Sinusoidal Data IV



$$||r||^2 = 13.0578, ||p_2^{LM}|| = 0.566443$$

$$x = (16.5319, 0.465955, 10.8305, 76.3247)$$

Method Comparisons

Exponential data (3 iterations):

- $||r||^2 = 6.01308$ with GN
- $||r||^2 = 6.01308$ with LMA

Sinusoidal data (2 iterations):

- $||r||^2 = 13.6556$ with GN
- $||r||^2 = 13.0578$ with LMA

Limitations

- Both GN and LMA approximate $\nabla^2 f(x)$ by eliminating the second term involving $\nabla^2 r$.
- If the residual is large or the model does not fit the function well, other methods must be used.
- Local minimum vs. global minimum

Bibliography I

- "Average Weather for Baton Rouge, LA - Temperature and Precipitation." The Weather Channel. 28 June 2012 (http://www.weather.com/weather/wxclimatology/monthly/graph/USLA0033)
- Gill, Philip E.; Murray, Walter. *Algorithms for the solution of the nonlinear least-squares problem*. SIAM Journal on Numerical Analysis 15 (5): 977-992. 1978.
- Griva, Igor; Nash, Stephen; Sofer Ariela. *Linear and Nonlinear Optimization*. 2nd ed. Society for Industrial Mathematics. 2008.
- Nocedal, Jorge; Wright, Steven J. *Numerical Optimization*, 2nd Edition. Springer, Berlin, 2006.

Bibliography II

"The Population of the United States." University of Illinois. 28 June 2012 (mste.illinois.edu/malcz/ExpFit/data.html).

Ranganathan, Ananth. "The Levenberg-Marquardt Algorithm." Honda Research Institute, USA. 8 June 2004. 1 July 2012 (http://ananth.in/Notes_files/lmtut.pdf).

Acknowledgements

- Dr. Mark Davidson
- Dr. Humberto Munoz
- Ladorian Latin