Inverse, Shifted Inverse, and Rayleigh Quotient Iteration as Newton’s Method

Richard Tapia
(Research joint with John Dennis)
Rice University

West Coast Optimization Meeting
May 4, 2013

- Talk and paper (in preparation) dedicated to Tony Chan on the occasion of his 60th Birthday
- A portion of this material was presented at the inaugural David Blackwell and Richard Tapia Conference at Cornell University in May of 2000 under the title “**If It Is Fast And Effective, It Must Be Newton’s Method**”

Preface

The Newton's Method Identification Problem

Given an iterative procedure that is effective and fast, and is certainly not know to be Newton's method, demonstrate that it is really a form of Newton's method by exhibiting the fundamental underlying (often well-hidden) nonlinear equation.

Our Message

"All" effective and fast methods are forms (perhaps very disguised) of Newton's Method. Moreover, some effective non-fast algorithms may also be forms on Newton's method.

Outline

- Preliminaries
- Examples of the Newton Identification Problem
- Development of algorithms for the eigenvalue problem
- Development of optimization algorithms and equivalences to eigenvalue algorithms
- An alternative development of Rayleigh Quotient Iteration
- Summary Statements

Cookbook Theory

Consider the fixed point problem

$$x^* = S(x^*)$$

and the obvious iterative method

$$x_{k+1} = S(x_k)$$

Def: Local convergence

$$x_k \longrightarrow x^* \quad \text{for all } \; x_0 \in N(x^*)$$

Cookbook Theory (cont.)

Def: Convergence rate

$$\left\|x_{k+1} - x^*\right\| \leq C\left\|x_k - x^*\right\|^p$$

Linear: $C < 1 \text{ and } p = 1$

Quadratic: $p = 2$

Cubic: $p = 3$

Theory in a Nutshell

Local convergence if

$$\rho\left(S'(x^*)\right) < 1$$

Quadratic convergence if

$$S'(x^*) = 0$$

Cubic convergence if

$$S'(x^*) = 0 \quad \text{and} \quad S''(x^*) = 0$$

$$\left[\left\| S''(x^*)\left(x_k - x^*, x_k - x^*\right)\right\| = 0\left(\left\|x_k - x^*\right\|^3\right)\right]$$

Newton's Method for $F(x)=0$

$$S(x) = x - F'(x)^{-1} F(x)$$

and

$$S'(x^*) = 0$$

Therefore: Newton's method is locally and quadratically convergent, i.e., it is effective and fast.

Application of the Newton Identification Approach to Several Examples from the Literature

The Babylonian’s Computation of $\sqrt{2}$

Consider: $x^2 = 2$

Write: $x = \dfrac{2}{x}$

Iterate: $x_{k+1} = \dfrac{2}{x_k}$

Failure:

$$x_1 = \frac{2}{x_0}$$

$$x_2 = x_0$$

$$x_3 = \frac{2}{x_0}$$

$$x_4 = x_0$$

The Babylonian's (cont.)

Prune and tune: add x to both sides

$$x + x = x + \frac{2}{x}$$

Simplify:

$$x = \frac{1}{2}\left(x + \frac{2}{x} \right)$$

Iterate:

$$x_0 = 1.00000$$
$$x_1 = 1.50000$$
$$x_2 = 1.41666$$
$$x_3 = 1.41421$$
$$x_4 = 1.41421$$

The resulting algorithm is effective and fast.

The Babylonian’s Method is Newton’s Method

$$F(x) = x^2 - 2$$

$$F'(x) = 2x$$

$$x_{k+1} = x_k - \frac{F(x_k)}{F'(x_k)}$$

$$x_{k+1} = x_k - \frac{x_k^2 - 2}{2x_k} = \frac{1}{2}\left(x_k + \frac{2}{x_k}\right)$$

Unconstrained Optimization

Problem: $\min_{x} f(x)$

Algorithm: $x_{k+1} = x_k^{\;x} + \Delta x_k$

where Δx_k solves the quadratic program

$$\min_{\Delta x} \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x$$

Unconstrained Optimization (cont.)

Experience: This algorithm is fast and effective.

Fact: This algorithm is Newton’s Method where the “equivalent” square nonlinear system is

$$F(x) \equiv \nabla f(x) = 0$$

Equality-Constrained Optimization

Problem:

$$\begin{aligned} &\min f(x) \\ &s.t. \;\; h(x) = 0 \end{aligned}$$

Idea: Combining Newton nonlinear equation and Newton unconstrained optimization approaches suggests

Equality-Constrained (cont.)

Algorithm: $x_{k+1} = x_k + \Delta x_k$

where Δx_k solves the quadratic program

$$\begin{aligned} \min_{\Delta x} \quad & \nabla f(x_k)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x_k) \Delta x \\ \text{s.t.} \quad & \nabla h(x_k) \Delta x + h(x_k) = 0 \end{aligned}$$

Fact: While this approach can be found in the literature it is not effective, i.e., does not give local convergence.

The Tuning and Pruning Process

Researchers led by MJD Powell (circa 1970) realized that “constraint curvature” had to be added to the model subproblem and converged upon the following so-called **Successive Quadratic Programming Method**

$$x_{k+1} = x_k + \Delta x_k$$

The Tuning and Pruning (cont.)

where Δx_k solves the quadratic program

$$\min_{\Delta x} \ \nabla f(x_k)^T \Delta x + \frac{1}{2} \Delta x^T \nabla_x{}^2 l(x_k, y_k) \Delta x$$
$$\text{s.t.} \quad \nabla h(x_k) \Delta x + h(x_k) = 0$$

Here $\nabla_x^2 l(x, y)$ is the Hessian with respect to x of the Lagrangian function

$$l(x, y) = f(x) + y^T h(x)$$

The Tuning and Pruning (cont.)

The y_k' s are updated as the multipliers obtained for the constraints in the solution of the quadratic program subproblem.

Experience: The SQP algorithm is effective and fast.

Fact: The SQP algorithm is Newton' s method where the "equivalent" square nonlinear system is

$$F(x,y) \equiv \nabla l(x,y) = 0$$

The Tuning and Pruning (cont.)

i.e.

$$\nabla_x l(x, y) = 0$$

$$h(x) = 0,$$

the first-order necessary conditions.

OK enough warming-up. Let’s get after the big score.

Rayleigh Quotient Iteration as Newton’s Method

The Quintessential “Tuning and Pruning” Process

The Symmetric Eigenvalue Problem

Given symmetric

$$A \in R^{nxn}$$

find

$$(x_*, \sigma_*)$$

such that

$$Ax_* = \sigma_* x_*$$

Rayleigh quotient:

$$\sigma_R(x) = \frac{x^T A x}{x^T x} \quad (x \neq 0)$$

Remark:

$$\sigma_R(\alpha x) = \sigma_R(x)$$

Chronology

- 1870's – Lord Rayleigh:

$$\left(A - \sigma_R(x) I\right) x_+ = e_1$$

(e_1 is the first unit coordinate vector)

1944 – H. Wielandt: Suggests fractional iteration (shifted and normalized inverse iteration)

$$\left(A - \sigma I\right)\hat{x} = x \qquad x_+ = \frac{\hat{x}}{\hat{x}_j}$$

Note: $\sigma = 0$ gives inverse power.

Chronology (cont.)

- 1945 – J. Todd (lecture notes), "approximately solve":

$$(A - \sigma_R(x) I) x_+ = 0$$

Chronology (cont.)

- 1949 – W. Kohn: (in a letter to the editor) suggests

$$(A-\sigma_R(x)I)x_+ = e_i$$

(any unit coordinate vector)

Kohn argues (without a rigorous proof) quadratic rate for $\{\sigma_k\}$.

Chronology (cont.)

- 1951 – S. Crandall:

$$\left(A - \sigma_R(x)I\right)x_+ = x$$

(this is unnormalized *RQI*)

Using an argument that tacitly assumed convergence, Crandall established <u>cubic</u> rate for $\{x_k\}$.

Remark: $\{x_k\}$ will not converge. However, $\sigma_R(x_k)$ can converge to σ_* .

Chronology (cont.)

- 1957 – A.M. Ostrowski:

  a) $(A-\sigma_R(x)I)x_+ = \eta \quad (\eta \neq 0)$ (viewed as implementation of Todd)
  Rigorously established quadratic rate for $\{\sigma_k\}$.

  b) Observed that Wielandt's fractional iteration combined with a) above leads to

  $$(A-\sigma_R(x)I)x_+ = x$$

  (unnormalized *RQI*)

  Ostrowski rigorously established cubic rate for $\{\sigma_k\}$.

  c) Ostrowski aware of Kohn, but unaware of Crandall, Forsythe points out (while paper in press) (10/57) Crandall's earlier work (*RQI* algorithm and cubic rate). Ostrowski acknowledges and remains *cool*.

Chronology (cont.)

- 1958 – A.M. Ostrowski (Ostrowski fights back):
  a) Points out cubic in x does not imply cubic in σ.
  b) Points out that a normalization* can be applied to x_+ to guarantee convergence of

$$\left\{ \frac{x_k}{\|x_k\|} \right\}$$ to an eigenvector.

* Wielandt had a normalization (∞ - norm)

Chronology (cont.)

This “tuning and pruning” process has taken us in the time period 1870’s to 1958 to

Rayleigh Quotient Iteration (*RQI*):

a) Solve: $\left(A-\sigma_R(x)I\right)y_+ = x$

b) Normalize: $x_+ = \dfrac{y_+}{\|y_+\|}$

Attributes:

a) Requires solution of linear system

b) Fast convergence (cubic)

Observation: At solution *RQI* is singular.

Some Points to Ponder

- Wielandt gave the eigenvalue approximation update formula

$$\sigma_+ = \sigma + \frac{1}{x_j}$$

Theorem. Wielandt iteration is equivalent to Newton's method on

$$Ax - \sigma x = 0$$

$$e_j^T x - 1 = 0$$

Some Points to Ponder (cont.)

- Wielandt stressed nonsymmetric A.
- Wielandt gave other eigenvalue formulas, but he never suggested the Rayleigh quotient. If he had he would have given normalized RQI in 1944.
- RQI performs better than Wielandt iteration for both symmetric and nonsymmetric matrices with real eigenvalues, except for the example matrix given in Wielandt’s paper.

Some Points to Ponder (cont.)

DID HELMUT WEILANDT KNOW OF THE RAYLEIGH QUOTIENT?

Some Points to Ponder (cont.)

- Peters and Wilkinson (1979) observe that x_+ obtained from inverse iteration and x_+ obtained from one step of Newton on

$$Ax - \sigma x = 0$$
$$x^T x - 1 = 0$$

are the same (Newton to first order)

Some Points to Ponder (cont.)

They propose Newton on

$$
\begin{aligned}
Ax - \sigma x &= 0 \\
e_j^T x - 1 &= 0.
\end{aligned}
$$

In this case the (∞-norm) normalization is automatically satisfied and the quadratic convergence is assured, unlike the two-norm case where the normalization may destroy the quadratic convergence. They did not realize that they had reinvented Wielandt iteration. But wait, for there is more.

Normalized Newton's Method: (NNM)

For the nonlinear equation problem

$$F(x) = 0$$

when it is known that $\|x_*\| = 1$.

a) Solve: $$F'(x)\Delta x = -F(x)$$

b) Normalize: $$x_+ = \frac{x + \Delta x}{\|x + \Delta x\|}$$

Attributes:

a) Requires solution of a linear system

b) Fast convergence (quadratic)

“BELIEF IN FOLKLORE”

There exists F such that $RQI \Leftrightarrow NNM$ on F.

- Our objective: Search for the missing F.
- Possible rewards:
  a) Explain why asymptotic singularity does not hurt phenomenal convergence rate of RQI.
  b) Explain why convergence is cubic and not just quadratic.

Historical Notes

- 1965 etc. – J. Wilkinson
- Instrumental in <u>exposing myth</u> that proximity of $A - \sigma_R(x)I$ to a singular matrix spoils the powerful convergence in practice (theory versus practice)

  (Angle not the magnitude that is important)
- 1969 – W. Kahan – B. Parlett

  a) Global convergence of RQI

  b) Monotone decreasing residuals

$$r = \left\| Ax - \sigma_x(x)x \right\|_2 \downarrow 0$$

Motivation (John Dennis Question)

$$\text{Is } RQI \Leftrightarrow \textit{Normalized Newton}$$

on

(*) $$F(x) \equiv \left(A - \sigma_R(x)I\right)x = 0 \qquad ?$$

Theorem: Given any x, not an eigenvector of A, the Newton iterate on (*) is ZERO.

(Normalized Newton is not defined)
(Newton's Method goes directly to the unique singular point of F.)

Nonlinear Programming
and
Our Equivalences

Equality Constrained Nonlinear Program

$$\begin{array}{l}\text{minimize } f(x) \\ \text{subject to } h(x) = 0\end{array}$$

$$f : R^n \longrightarrow R \quad \text{and} \quad h : R^n \longrightarrow R^m (m < n)$$

Lagrangian Function:

$$1(x, \lambda) = f(x) + \lambda^T h(x)$$

$$1 : R^n \times R^m \longrightarrow R$$

Lagrange Multiplier Rule (First Order Necessary Conditions)

$$\nabla \ell(x,\lambda) = \begin{pmatrix} \nabla_x \ell(x,\lambda) \\ h(x) \end{pmatrix} = 0.$$

i.e.

$$\begin{aligned} \nabla f(x) + \lambda_1 \nabla h_1(x) + \ldots + \lambda_m \nabla h_m(x) &= 0 \\ h_1(x) &= 0 \\ \ldots \\ h_m(x) &= 0 \end{aligned} \qquad (*)$$

Constructing The Eigen-Nonlinear Program

Consider $$\min_{x \neq 0} \frac{x^T A x}{x^T x}$$

This problem is not well posed for Newton's Method. It has several continua of solutions.

We can obtain isolated solutions for simple eigenvalues by adding the constraint

$$x^T x = 1.$$

The Eigen-Nonlinear Program

$$\min_{x} \; x^T A x$$

$$\text{s.t. } (x^T x - 1) = 0$$

Necessary conditions:

$$(A - \lambda I)x = 0$$

$$(x^T x - 1) = 0$$

Observations:

Unit Eigenvectors ⇔ Stationary Points

Eigenvalues ⇔ Lagrange Multipliers

The Mathematicians Crutch

Reduce the problem at hand to a solved problem, or a sequence of such solved problems.

Example: Differentiation and the partial derivative

Example: Newton's Method – reduces the solution of a square nonlinear system of equations, to solving a sequence of square linear systems of equations.

The present case: SUMT– Sequential unconstrained minimization techniques. We reduce the solution of the constrained optimization problem to solving a sequence of unconstrained problems.

Penalty Function Method (Courant 1944)

$$\min_{x} \quad P(x;\rho) = f(x) + \frac{\rho}{2} h(x)^T h(x) \quad (\rho > 0)$$

Observation: Stationary points of our equality constrained optimization problem do not correspond to stationary points of $P(x;\rho)$

Proof:

$$\left\{ \begin{array}{r} \nabla f(x) + \nabla h(x)\lambda = 0 \\ h(x) = 0 \end{array} \right\} \begin{array}{c} \nLeftarrow \\ \nRightarrow \end{array} \left\{ \nabla f(x) + \nabla h(x) \rho h(x) = 0 \right\}$$

Penalty Function Method (cont.)

Observation: We need $\rho h(x)=\lambda$ at a solution but $\lambda \neq 0$ and $h(x)=0$; hence we must have

$$\rho = +\infty \qquad (\infty \cdot 0 = \lambda)$$

We therefore perform the minimization sequentially for a sequence $\{\rho_k\}$ such that $\rho_k \longrightarrow +\infty$.

Equivalence Result

Normalized Inverse Iteration

a) Solve $Ay = x$

b) Normalize $x_+ = \frac{y}{\|y\|}$

Theorem: Normalized inverse iteration is equivalent to normalized Newton's method on the penalty function for the eigen-nonlinear program

$$P(x;\rho) = x^T Ax + \frac{\rho}{2}(x^T x - 1)^2 \text{ for any } \rho \neq 0.$$

Equivalence Result (cont.)

But wait, we don't get quadratic convergence. Why?

Answer: $\nabla_x P(x;\rho) = 2Ax + \rho(x^T x - 1)x$

For eigenpair (x^*, σ^*) with $\sigma^* \neq 0$ and $\|x^*\| = 1$
we have $\nabla_x P(x^*, \rho) = 2Ax^* = 2\sigma^* x^* \neq 0$

Therefore, the unit eigenvector is <u>not</u> a stationary point of $P(x;\rho)$ for any *ρ*, as we pointed out before. The normalization forces convergence but not to a stationary point.

Multiplier Method (Hestenes/Powell) (1968-1969)

Augmented Lagrangian:

$$L(x,\lambda;\rho) = f(x) + \lambda^T h(x) + \frac{1}{2}\rho h(x)^T h(x) \quad (\rho > 0)$$

Observation: As before stationary points of our equality constrained optimization problem do not correspond to stationary points (in x) of $L(x,\lambda;\rho)$.

Multiplier Method (cont.)

Multiplier Method:

Given (x,λ,ρ) obtain x_+ as the solution of

$$\min_x L(x,\lambda;\rho)$$

Let $\lambda \longrightarrow \lambda_+$ and $\rho \longrightarrow \rho_+$

Remark: The multiplier method does not need $\rho \longrightarrow +\infty$ as did the penalty function method. In the penalty function method $\rho \rightarrow \infty$ causes arbitrarily bad conditioning.

Equivalence Result

Normalized Shifted Inverse Iteration with shift σ

a) Solve $(A - \sigma I)y = x$

b) Normalize $x_{+} = \dfrac{y}{\|y\|}$

Theorem: Normalized shifted inverse iteration with shift σ is equivalent to normalized Newton's method on the augmented Lagrangian for the eigen-nonlinear program with multiplier estimate σ, i.e., on

$$L(x;\sigma;\rho) = x^T Ax + \sigma(x^T x - 1) + \frac{\rho}{2}(x^T x - 1)^2 \quad \text{for any } \rho \neq 0$$

Equivalence Result (cont.)

But wait, we don't have quadratic convergence. Why?

Answer: Same as for the penalty function case. The normalization forces convergence, but not to a stationary point, but to a solution of our problem.

In the Interest of Efficiency We Ask

Question: Can we move away from a sequence of unconstrained optimization problems to just one unconstrained optimization problem, i.e., a so-called exact penalty function?

The Fletcher Exact Penalty Function (1970)

$$\min_{x} \ \phi(x,\rho)$$

where $\phi(x,\rho) = L(x,\lambda(x);\rho)$

and $\lambda(x)$ is the well-known Lagrange Multiplier Approximation Formula

$$\lambda(x) \equiv -\left(\nabla h(x)^T \nabla h(x)\right)^{-1} \nabla h(x)^T \nabla f(x)$$

observe that $\lambda(x^*) = \lambda^*$.

Theorem:

For ρ sufficiently large our constrained minimizers are minimizers of ϕ, i.e., ϕ is an exact penalty function.

Criticism:

- How large is sufficiently large
- Derivatives are out of phase, i.e.,

  $\nabla \phi(x)$ involves $\nabla^2 f(x), \nabla^2 h_i(x)$

Correcting the Deficiencies of the Fletcher Exact Penalty Function

The Multiplier Substitution Equation (Tapia 1978)

Instead of considering the unconstrained optimization problem

$$\min_{x} \ \varphi(x;\rho)$$

consider the nonlinear equation problem

$$F(x;\rho) = \nabla_x L(x,\lambda(x);\rho) = 0 \qquad \rho \neq 0$$

Observation:

$$F(x;\rho) = \nabla_x l(x, \lambda_\rho(x))$$

where $$\lambda_\rho(x) = \lambda(x) + \rho h(x).$$

Remark: System is a square nonlinear system of equations; hence, the use of Newton's method is appropriate.

Theorem

Stationary points of the equality constrained optimization problem correspond to zeros of the multiplier substitution equation.

Proof (very pretty)

$$\nabla f(x) + \nabla h(x)\big(\lambda(x) + \rho h(x)\big) = 0$$

Substitute $\lambda(x)$

$$\nabla f(x) - \nabla h(x)\big(\nabla h(x)^T \nabla h(x)\big)^{-1} \nabla h(x)^T \nabla f(x) +$$
$$\nabla h(x)\rho h(x) = 0$$

Theorem (cont.)

multiply by $\nabla h(x)^T$ to obtain

$$\nabla h(x)^T \nabla h(x) \rho h(x) = 0$$

therefore $\rho h(x) = 0$ and $\rho \neq 0$

so $h(x) = 0$ and $\nabla f(x) + \nabla h(x) \sigma(x) = 0$

Equivalence Result

Normalized Rayleigh Quotient Iteration

a) Solve $(A - \sigma_R(x)I)y = x$

b) Normalize $x_+ = \frac{y}{\|y\|}$

Theorem: Normalized Rayleigh Quotient Iteration is equivalent to normalized Newton's method on the multiplier substitution equation, for any $\rho \neq 0$, i.e., on

$$F(x;\rho) = \nabla_x L(x, \lambda(x);\rho) = 0$$

Equivalence Result (cont.)

where

$$\lambda(x) = -(\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x).$$

Remark: We finally have our equivalence.

But wait, we don’t have quadratic convergence (indeed we have cubic). Why?

Some Observations on the Cubic Convergence of RQI

Theorem

If σ_i is simple, then $F'(x;\rho)$ is nonsingular for all x in a neighborhood of each unit eigenvector corresponding to σ_i and for all $\rho \neq 0$.

Moreover, the Normalized Newton Method is locally cubically convergent with convergence constant less than or equal to one.

Implication

Although asymptotically RQI becomes a singular algorithm, this singularity is removable.*

**Jim Wilkinson in the 1960’s and 1970’s repeatedly stated that the singularity doesn’t hurt – the angle is good, the magnitude is not important.*

Cubic Convergence

Newton Operator: $N(x) = x - F'(x)^{-1} F(x)$

Normalized Newton Operator: $S(x) = \dfrac{N(x)}{\|N(x)\|_2}$

Lemma: Let x_* be a unit eigenvector corresponding to a simple eigenvalue, then

a. $S'(x_*) = 0$ (quadratic)

b. $\|S''(x_*)(x - x_*, x - x_*)\| \leq K \|x - x_*\| [x_*^T (x - x_*)]$

Cubic Convergence (cont.)

Surprise: If $\|x_*\| = \|x\| = 1,$ then

$$x_*^T(x_* - x) = 1 - x_*^T x = \frac{1}{2}\left(1 - 2x_*^T x + 1\right) = \frac{1}{2}\|x_* - x\|^2$$

Conclusion:

$$\| S''(x_*)(x - x_*, x - x_*) \| = 0\left(\|x - x_*\|^3 \right)$$

Must normalize in 2-norm for good x-behavior.

Two Fascinating Reflections

- The beauty of Euclidean geometry

If $\|x\| = \|y\| = 1$ then

$$x^T(x-y) = \frac{1}{2}\|x-y\|^2$$

Two Fascinating Reflections (cont.)

- The Babylonian shift (a universal tool?)
  - In “deriving” Newton’s Method for $x^2 = 2$, when $S(x) = \frac{2}{x}$ uniformly failed they corrected by shifting by x, i.e., they considered

$$S_{new}(x) = \frac{1}{2}(S(x) + x)$$

Two Fascinating Reflections (cont.)

- In deriving Newton's method as Rayleigh quotient iteration for symmetric eigenvalue problem, when $\lambda(x) = \lambda_R(x) = \dfrac{x^T Ax}{x^T x}$

  uniformly failed, we considered

  $$\lambda(x) = \lambda_R(x) + \frac{1}{2}\left(x^T x - 1\right)$$

  so that

  $$\nabla\lambda(x) = \nabla\lambda_R(x) + x$$

  i.e., we corrected by shifting by x.

An Alternative Approach to Deriving RQI

Consider the eigen-nonlinear program

$$\begin{aligned} \min_{x} \quad & x^T A x \\ \text{s.t.} \quad & (1 - x^T x) = 0 \end{aligned}$$

Newton's Method

$$x_+ = x + \Delta x$$

$$\lambda_+ = \lambda + \Delta\lambda$$

where

$$\begin{pmatrix} A - \lambda I & -x \\ -x^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} (A - \lambda I)x \\ \frac{1}{2}(1 - x^T x) \end{pmatrix}$$

convergence of $\{(x_k, \lambda_k)\}$ is quadratic.

Normalized Newton’ s Method (closing the norm constraint)

$$\lambda_+ = \lambda + \Delta\lambda$$

$$x_+ = \frac{(x+\Delta x)}{\|x+\Delta x\|_2}$$

Tapia- Whitley 1988:

The convergence rate of $\{x_k\}$ is $1+\sqrt{2}$
The convergence rate of $\{\lambda_k\}$ is $1+\sqrt{2}$

Closing the constraint improved the convergence rate from 2 to $1+\sqrt{2}$.

Normalized Newton’s Method (cont.)

Remark: Recall that Peters and Wilkinson were concerned that this two-norm normalization would destroy quadratic convergence. Indeed, it improves the rate.

Normalized Newton’s Method (cont.)

Idea: Let’s choose λ_+ to “close” the equation

$$(A - \lambda I)x_+ = 0$$

Remark: We can’t do this, because for fixed x_+ this is an overdetermined system in λ .

O.K., let’s “close” in the least-squares sense, i.e., let λ_+ be the solution of

$$\min_{\lambda} \quad \|(A - \lambda I)x_+\|_2$$

Normalized Newton’s Method (cont.)

this gives
$$\lambda_+ = \frac{x_+^T A x_+}{x_+^T x_+} = \sigma_R(x_+)$$

Our closed in x and least-squares closed in λ Newton’s method becomes

$$x_+ = \frac{x + \Delta x}{\left\| x + \Delta x \right\|_2}$$

$$\lambda_+ = \frac{x_+ A x_+}{x_+^T x_+}$$

Normalized Newton’s Method (cont.)

Theorem: This algorithm is RQI. Therefore

convergence of $\{x_k\}$ is cubic

convergence of $\{\lambda_k\}$ is cubic.

Summary

- Newton’s Method gives quadratic convergence.
- Newton’s Method followed by (normalization) a closing of the norm constraint gives a convergence rate of $1+\sqrt{2}$.
- Newton’s Method followed by a (normalization) closing of the norm constraint, followed by a least-squares closing of the gradient of the Lagrangian equal to zero equation, gives a convergence rate of 3 and also gives us Rayleigh Quotient Iteration.