On double Newton steps

Tobias Damm
Institute of Applied Mathematics
Department of Numerical Mathematics
TU Braunschweig
Germany
t.damm@tu-bs.de

Abstract

In this note we adapt some results of C-H. Guo and P. Lancaster on the use of double Newton steps to a more general situation. Numerical examples illustrate the idea.

1 Introduction

In a series of papers [8, 6, 7], C.-H. Guo and P. Lancaster have established the use of double Newton steps in the iterative solution of Riccati equations $\mathcal{R}(X) = 0$, whose linearization is singular at the limit point X_+, i.e. $\det \mathcal{R}'_{X_+} = 0$. This situation occurs e.g. in robust control or disturbance attenuation problems, where a Riccati equation has to be solved for a maximal robustness margin or a minimal attenuation value. While general results (e.g. [10]) still guarantee that a Newton sequence (X_k) converges to X_+ if $\sigma(\mathcal{R}'_{X_0}) \subset \mathbb{C}_-$, it is well-known that the convergence cannot be expected to be quadratic in this case. But, as Guo and Lancaster point out, if the convergence is only linear, it can be of advantage to consider a double Newton step in each iteration. Starting from a given iterate of the Newton iteration, a double Newton step eventually yiels a satisfactory solution at once, long before the standard Newton procedure converges. If, however, the approximation produced by the double step is not acceptable yet, it is discarded, and the process continues with a single Newton step.
A similar situation arises in H^∞-type control problems for stochastic systems. Here one has to deal with generalized Riccati-type matrix equations. In [12, 4] non-local convergence results for Newton's method applied to these generalized matrix

equations have been proven. Like in the case of standard Riccati equations one can prove the convergence of the Newton sequence (X_k) to X_+ under the assumption $\sigma(\mathcal{R}'_{X_0}) \subset \mathbb{C}_-$, even if $\det \mathcal{R}'_{X_+} = 0$. It is natural to ask whether the results on double Newton steps can be transferred to this situation as well. It is the object of this note to answer this question. While the theorems in [8, 6, 7] cannot be applied directly, it only takes slight modifications of the ideas. To develop the central points clearly, we use the abstract set-up of a Banach space. The context of Riccati equations is important insofar, as we assume some a priori knowledge on the convergence of the Newton sequence.

2 The use of double Newton steps

In the following let $f : \operatorname{dom} f \subset X \to X$ be a Fréchet differentiable mapping on a convex open domain $\operatorname{dom} f$ in a Banach space X. By ϕ_x we denote the second order remainder term in the Taylor expansion of f at x, such that for $h \in X$

$$f(x+h) \quad = \quad f(x) + f'_x(h) + \phi_x(h) \; .$$

Let there exist numbers $L_0, L_1 \geq 0$ such that

$$\|\phi_x(x-y)\| \quad \leq \quad L_0\|x-y\|^2 \quad \text{for all} \;\; x, y \in \operatorname{dom} f \; ; \tag{1}$$

and

$$\|(\phi_x - \phi_y)(x-y)\| \quad \leq \quad L_1\|x-y\|^3 \quad \text{for all} \;\; x, y \in \operatorname{dom} f \; . \tag{2}$$

These conditions are satisfied, for instance, if the third Fréchet derivative of f exists on $\operatorname{dom} f$ and is bounded.
We assume that $f(x_+) = 0$ for some $x_+ \in \operatorname{dom} f$ and that there exists a Newton sequence $x_0, x_1, \ldots \in \operatorname{dom} f$, converging to x_+. The following technical lemma is a variation of a result in [8]. We need it to describe the case, when the Newton sequence is not quadratically convergent.

Lemma 2.1 *Let $\theta > 0$ and set*

$$Q_\theta \quad = \quad \Big\{ k \in \mathbb{N} \mid \|f'_{x_+}(x_{k+1} - x_+)\| \geq \theta \|x_{k+1} - x_+\| \Big\} \; .$$

Then for all numbers $\kappa > \frac{L_0}{\theta}$ there exists an index k_0, such that

$$\|x_{k+1} - x_+\| \quad \leq \quad \kappa \|x_k - x_+\|^2$$

for all $k \in Q_\theta$ with $k \geq k_0$.

Proof: Let $k \in Q_\theta$. Since $f(x_+) = 0$ we have the Taylor expansion

$$f(x_{k+1}) \;=\; f'_{x_+}(x_{k+1} - x_+) + \phi_{x_+}(x_{k+1} - x_+)\ .$$

We apply the triangle inequality and the Lipschitz condition to obtain

$$\begin{aligned}\|f(x_{k+1})\| &\geq \theta\|x_{k+1} - x_+\| - \|\phi_{x_+}(x_{k+1} - x_+)\| \\ &\geq (\theta - L_0\|x_{k+1} - x_+\|)\|x_{k+1} - x_+\|\ .\end{aligned}$$

On the other hand we have for the Newton iterates

$$\begin{aligned}\|f(x_{k+1})\| &= \|\phi_{x_{k+1}}(x_{k+1} - x_k)\| \\ &\leq L_0\|(x_{k+1} - x_+) + (x_+ - x_k)\|^2 \\ &\leq L_0\|x_{k+1} - x_+\|^2 + 2L_0\|x_{k+1} - x_+\|\,\|x_k - x_+\| \\ &\quad + L_0\|x_k - x_+\|^2\ .\end{aligned}$$

Combining the two inequalities for $\|f(x_{k+1})\|$ we finally get

$$\begin{aligned}\|x_k - x_+\|^2 &\geq \Big(\frac{\theta}{L_0} - 2\|x_{k+1} - x_+\| - 2\|x_k - x_+\|\Big)\|x_{k+1} - x_+\| \\ &\geq \frac{1}{\kappa}\|x_{k+1} - x_+\| \quad \text{for sufficiently large } k \in Q_\theta\ .\end{aligned}$$

This completes the proof. □

Obviously, if for some $\theta > 0$ the set $\mathbb{N}\backslash Q_\theta$ is finite, then convergence takes place quadratically with factor κ. If, however, f'_{x_+} is singular or almost singular, such a θ does not exist or it might be close to the machine precision. Hence for (almost) arbitrarily small $\theta > 0$ there exists an index k with

$$\|f'_{x_+}(x_k - x_+)\| \;<\; \theta\|x_k - x_+\|\ . \tag{3}$$

In such a situation a double Newton step can be useful. The key observation is, that $f(x_k)$ is almost equal to $\phi_{x_k}(x_+ - x_k)$. Namely, by the Taylor expansion at x_+ we have

$$\|f(x_k) - \phi_{x_k}(x_+ - x_k)\| \;=\; \|f'_{x_+}(x_k - x_+) + \phi_{x_+}(x_k - x_+) - \phi_{x_k}(x_+ - x_k)\|$$

The Taylor expansion at x_k therefore yields

$$\begin{aligned}\|2f(x_k) - f'_{x_k}(x_k - x_+)\| &= \|f(x_k) - \phi_{x_k}(x_+ - x_k)\| \\ &= \|f'_{x_+}(x_k - x_+) + \phi_{x_+}(x_k - x_+) - \phi_{x_k}(x_+ - x_k)\| \\ &\leq (\theta + L_1\|x_+ - x_k\|^2)\|x_+ - x_k\|\ ,\end{aligned} \tag{4}$$

which is close to zero for sufficiently small θ and $\|x_k - x_+\|$. Thus, the *double Newton step*

$$y_{k+1} = x_k - 2{f'}^{-1}_{x_k}(f(x_k)) \tag{5}$$

is likely to give a very good approximation to x_+ *at once*.
This argument of course requires also $\|f'^{-1}_{x_k} f'_{x_+}(x_k - x_+)\|$ to be small. To give a strict justification, we need some bound on the growth of $\|f'^{-1}_{x_k}\|$ as $k \to \infty$.

Lemma 2.2 *For some $c > 0$ and all sufficiently large $j \in \mathbb{N}$ let*

$$\|f'^{-1}_{x_j}\| < c\|x_j - x_+\|^{-1} . \tag{6}$$

If (3) holds for some $k \in \mathbb{N}$, then y_{k+1} from (5) satisfies $\|y_{k+1} - x_+\| \leq c\theta + cL_1\|x_k - x_+\|^2$.

Proof: Inserting (5) in (4) we have

$$\begin{aligned} \|y_{k+1} - x_+\| &= \left\| f'^{-1}_{x_k}\Big(f'_{x_k}(x_k - x_+) - 2f(x_k)\Big)\right\| \\ &\leq \|f'^{-1}_{x_k}\|\Big(\theta + L_1\|x_+ - x_k\|^2\Big)\|x_+ - x_k\| \\ &\leq c\theta + cL_1\|x_+ - x_k\|^2 , \end{aligned}$$

which we wanted to show. □

If cL_1 and $c\theta$ are small (e.g. $L_1 = 0$ for purely quadratic equations) then a double Newton step can increase the precision of the current iterate significantly. Since the computation of $f'^{-1}_{x_k}(f(x_k))$ is the most expensive operation, it is relatively cheap to try a double Newton step in each iteration. The algorithm can stop immediately if $\|f(y_{k+1})\|$ turns out to be small enough, while otherwise it continues with x_{k+1} obtained by a single Newton step.
Our results partially extends [8, Theorem 3.2] and [6, Theorem 5.2], since our set-up is more general. On the other hand, it must be stressed that in [8, 6] much stronger assertions have been obtained concerning the growth condition (6). That is, Guo and Lancaster were able to give an algebraic characterization of the case, when (6) is fulfilled. To achieve this in our general set-up is still an open problem to us.

3 Numerical examples

In the following we consider rational matrix equations, which occur in stochastic control. They are of the general form

$$\begin{aligned} \mathcal{G}(Y) &= -YA^* - AY - \sum_{j=1}^{N} YA_0^{j*}Y^{-1}A_0^jY + YP_0Y \\ &\quad - \Big(B + \sum_{j=1}^{N} YA_0^{j*}Y^{-1}B_0^j - YS_0\Big) \end{aligned} \tag{7}$$

$$\times \left(Q_0 - \sum_{j=1}^{N} B_0^{j*} Y^{-1} B_0^j\right)^{-1} \left(B + \sum_{j=1}^{N} B_0^{j*} Y^{-1} A_0^j Y - S_0 Y\right),$$

which will be specified in the examples. The task is to find the largest solution $Y_+ > 0$ of this equation. In [3] it was shown that – under given assumptions, which are satisfied in our examples – the Newton sequence

$$Y_{k+1} \quad = \quad Y_k - \mathcal{G}'^{-1}_{Y_k}(\mathcal{G}(Y_k))$$

converges to Y_+ (whenever such a solution exists), if $\sigma(\mathcal{G}'_{Y_0}) \subset \mathbb{C}_-$. Moreover, if $P_0 \leq 0$ and (A, P_0) is observable, then a matrix Y_0 satisfying $\sigma(\mathcal{G}'_{Y_0}) \subset \mathbb{C}_-$ can be found as $Y_0 = \nu Y_*$, where Y_* solves the standard Riccati equation

$$YA^* + AY + YP_0Y = I$$

and $\nu > 0$ is suffciently large. We will use this choice in our computations.
In each of our examples the matrix equation depends on a parameter. From the underlying problem it is clear that there exists a threshold value of this parameter, where the equation becomes unsolvable. If the parameter value is close to this threshold value, then the derivative $\mathcal{G}'_{Y_+}$ at the cooresponding solution Y_+ is almost singular. This situation will be studied. The threshold value is assumed to be known. It has been identified by a binary search.

3.1 A two-cart system

In [2] we have discussed the matrix equation

$$\begin{aligned}\mathcal{G}^\gamma(Y) \quad &= \quad -YA^* - AY - YA_0^*Y^{-1}A_0Y - YC^*CY \\ &\quad + B_2(D^*D)^{-1}B_2^* - \gamma^{-2}B_1B_1^* \\ &= \quad 0\end{aligned}$$

which corresponds to an H^∞-control problem for a two-cart system, introduced in [11]. The parameters are given by

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -5/4 & 5/4 & 0 & 0 \\ 5/4 & -5/4 & 0 & 0 \end{bmatrix}, \quad A_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1/4 & 1/4 & 0 & 0 \\ 1/4 & -1/4 & 0 & 0 \end{bmatrix},$$

$$\begin{bmatrix} B_1^T \\ B_2^T \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We are looking for a solution $Y > 0$. It can be shown that such a solution exists for $\gamma > 1.8923$. Moreover, the Newton sequence

$$Y_{k+1} \;=\; Y_k - (\mathcal{G}^{\gamma\,\prime}_{Y_k})^{-1}(\mathcal{G}^\gamma(Y_k))\;,$$

converges to the largest (with respect to the Loevner ordering) solution Y_+, whenever $\sigma(\mathcal{G}^{\gamma\,\prime}_{Y_0}) \subset \mathbb{C}_-$. For instance, we found the matrix

$$Y_0 \;=\; \begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 2 \end{bmatrix} \tag{8}$$

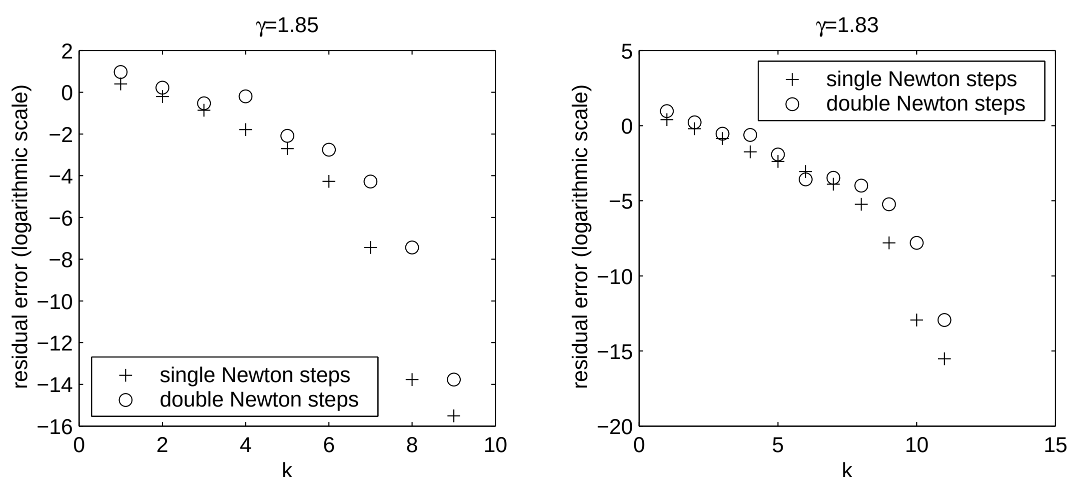
to satisfy this condition independently of γ.

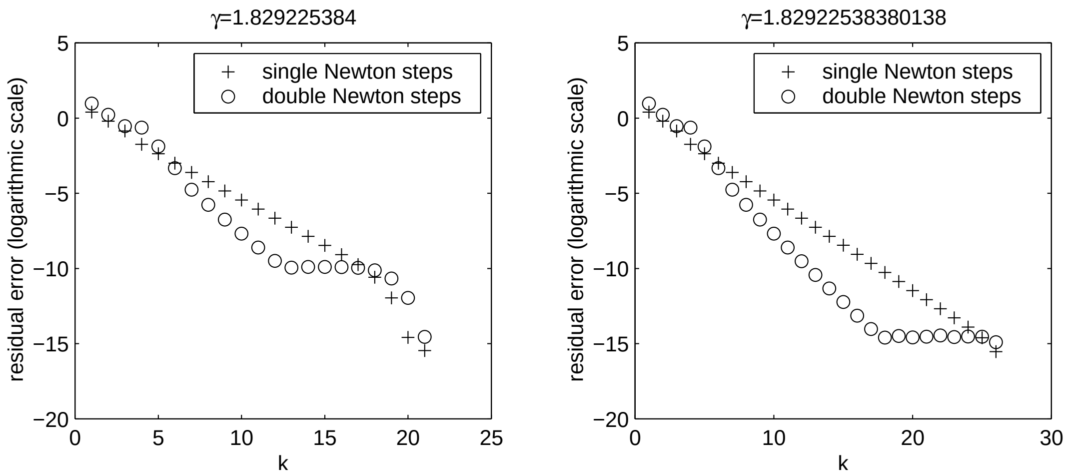
If γ is close to the optimal attenuation value $\gamma \approx 1.893$, then the Newton iteration is rather slow and almost linear. As we have pointed out in Section 2, we can benefit from a double Newton step in this situation.

To see this effect, we consider the cases $\gamma = 2$ and $\gamma \approx 1.893$ and apply Newton's method to the equation $\mathcal{G}^\gamma(Y) = 0$ with initial matrix Y_0 from (8). In each step of the iteration we compute, additionally, the matrix

$$Z_{k+1} \;=\; Y_k - (\mathcal{G}'_{Y_k})^{-1}(\mathcal{G}(Y_k))$$

and the residual errors $\|\mathcal{G}^\gamma(Y_{k+1})\|$ and $\|\mathcal{G}^\gamma(Z_{k+1})\|$. These residual errors are depicted in the following figures.





We see that for $\gamma = 1.85$ the Newton iteration converges quadratically and the double Newton steps do not bring any advantage. But if γ gets closer to the optimum, the situation changes. In the last figure, the convergence of the Newton iteration is practically linear and it takes about 25 steps to compute the solution with acceptable accuracy. The same accuracy, however, is obtained by a double Newton step already in the 18th iteration. If the Newton iteration is used to improve an aproximate solution, which is already close to the limit point, then a double Newtons step might yield a significant improvement.

3.2 A car-steering model

We consider the four-wheel car-steering model discussed in [1]. An important feature of this model is the uncertainty of the adhesion coefficient μ between tyre and road-surface. For instance, the values $\mu = 1$, $\mu = 0.5$, and $\mu = 0.15$ are assumed on dry road, on a wet road, and on ice, respectively. Here, we model the coefficient μ as a stochastic process with mean value 0.5 and intensity σ. Thus, the state-space system from [1] takes the stochastic form

$$dx \;=\; (Ax + B_2 u)\,dt + (A_0^{(1)}x + B_{20}^{(1)}u)\sigma\,dw_1 + (A_0^{(2)}x + B_{20}^{(2)}u)\sigma\,dw_2\;,$$

with

$$A = \begin{bmatrix} -\frac{c_r+c_f}{2mv} & \frac{c_r\ell_r - c_f\ell_f}{2mv^2} - 1 \\ \frac{c_r\ell_r+c_f\ell_f}{2J} & -\frac{c_r\ell_r^2+c_f\ell_f^2}{2Jv} \end{bmatrix},\; A_0^{(1)} = \begin{bmatrix} \frac{-c_f}{mv} & \frac{-c_f\ell_f}{mv^2} \\ \frac{c_f\ell_f}{J} & \frac{-c_f\ell_f^2}{Jv} \end{bmatrix},\; A_0^{(2)} = \begin{bmatrix} \frac{-c_r}{mv} & \frac{c_r\ell_r}{mv^2} \\ \frac{c_r\ell_r}{J} & \frac{-c_r\ell_r^2}{Jv} \end{bmatrix},$$

$$B = \begin{bmatrix} \frac{c_f}{2mv} & \frac{c_r}{2mv} \\ \frac{c_f\ell_f}{2J} & \frac{-c_r\ell_r}{2J} \end{bmatrix} \; B_{20}^{(1)} = \begin{bmatrix} \frac{c_f}{mv} & 0 \\ \frac{c_f\ell_f}{J} & 0 \end{bmatrix},\; B_{20}^{(2)} = \begin{bmatrix} 0 & \frac{c_r}{mv} \\ 0 & \frac{-c_r\ell_r}{J} \end{bmatrix}.$$

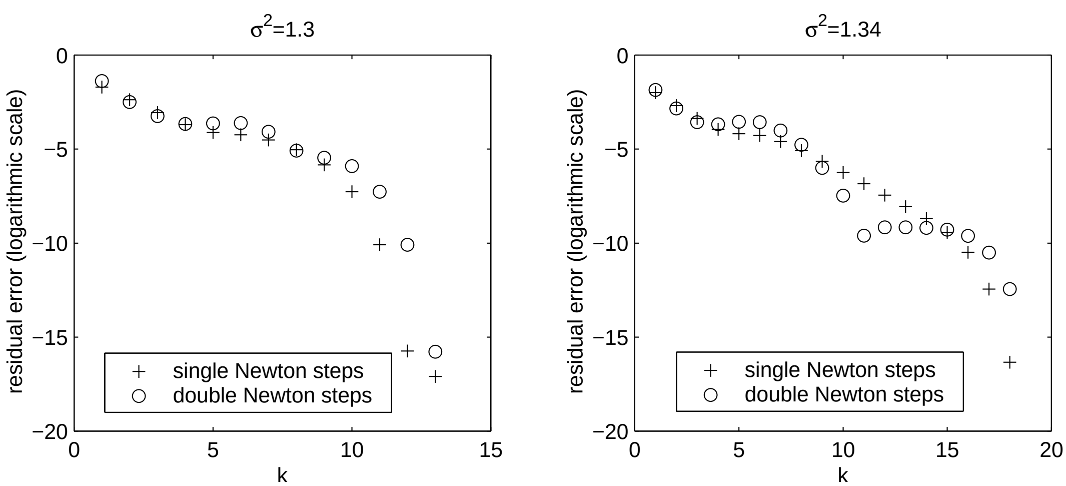
The state vector $x = [\beta, r]^T$ contains the sideslip angle β and the yaw rate r. The front and rear steering angles $[\delta_f, \delta_r]^T = u$ are the control variables. The parameter values are $\ell_f = 3.67\,[\mathrm{m}]$, $\ell_r = 1.93\,[\mathrm{m}]$, $c_f = 198000\,[\mathrm{N/rad}]$, $c_r = 470000\,[\mathrm{N/rad}]$, $v = 10[\mathrm{m/s}]$, $m = 10000[\mathrm{kg}]$, $J = i^2\mu m\,, i^2 = 10.85[\mathrm{m}^2]$.
We are looking for the maximal noise intensity σ for the system to be stabilizable in mean-square. From the theory, we know that the system is stabilizable, if and only if there exists a $Y > 0$, such that

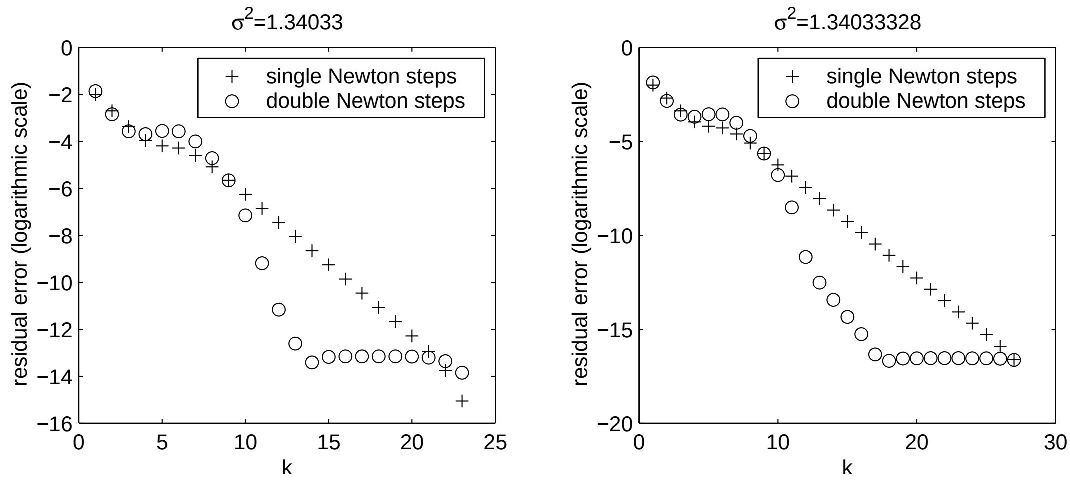
$$\begin{aligned}
\mathcal{G}^\sigma(Y) &= -YA^* - AY - \sigma^2 \sum_{j=1}^{2} YA_0^{j*}Y^{-1}A_0^jY - I \\
&\quad - YB\left(-\sigma^2 \sum_{j=1}^{2} B_0^{j*}Y^{-1}B_0^j - I\right)^{-1} B^*Y \\
&= 0\,.
\end{aligned}$$

A stabilizing feedback control $u = Fx$ is then given by

$$F = -\left(\sigma^2 \sum_{j=1}^{2} B_0^{j*}Y^{-1}B_0^j + I\right)^{-1} B^*$$

where Y_+ is the largest solution of $\mathcal{G}^\sigma(Y) = 0$.
The threshold value for the system to be stabilizable is approximately given by $\sigma^2 = 1.340327$. We visualize the residual errors $\|\mathcal{G}^\sigma(Y_k)\|$ and $\|\mathcal{G}^\sigma(Z_k)\|$ for the Newton iterates Y_k produced by single and Z_k produced by double Newton steps, when σ is close to this threshold value.





We observe the same effect as in the previous example. The closer σ is to the threshold value, the more we can take advantage from the double Newton step.

3.3 An automobile suspension system

A stochastic model for an automobile suspension system with uncertain stiffness parameter was introduced in [5]. In [3] we discussed the linearized state space-model given by

$$\begin{aligned} dx &= Ax\,dt + B_1 v\,dt + B_2 u\,dt + \sigma(A_0 x\,dw + B_{10} v\,dw)\,, \\ z &= Cx + D_2 u\,, \end{aligned}$$

where

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{-k_1}{m_1} & \frac{k_1}{m_1} & \frac{-c}{m_1} & \frac{c}{m_1} \\ \frac{k_1}{m_2} & \frac{-(k_1+k)}{m_2} & \frac{c}{m_2} & \frac{-c}{m_2} \end{bmatrix}, \quad A_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{-\sigma}{m_2} & 0 & 0 \end{bmatrix},$$

$$B_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{k}{m_2} \end{bmatrix}, \quad B_{10} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{\sigma}{m_2} \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 \\ 0 \\ \frac{b}{m_1} \\ 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0 \\ 0 \\ \varepsilon \end{bmatrix}.$$

Here $m_2 = 28.57$[kg], $k = 120$[kN/m], $k_1 = 20$[kN/m], $c = 1.6$[kNs/m], $m_1 = 1000$[kg], and $\epsilon = 0.01$ is a regularization parameter. The noise intensity σ is set to $\sigma = 0.2$. The task is to stabilize the system and to diminish the effect of the disturbance v on the output

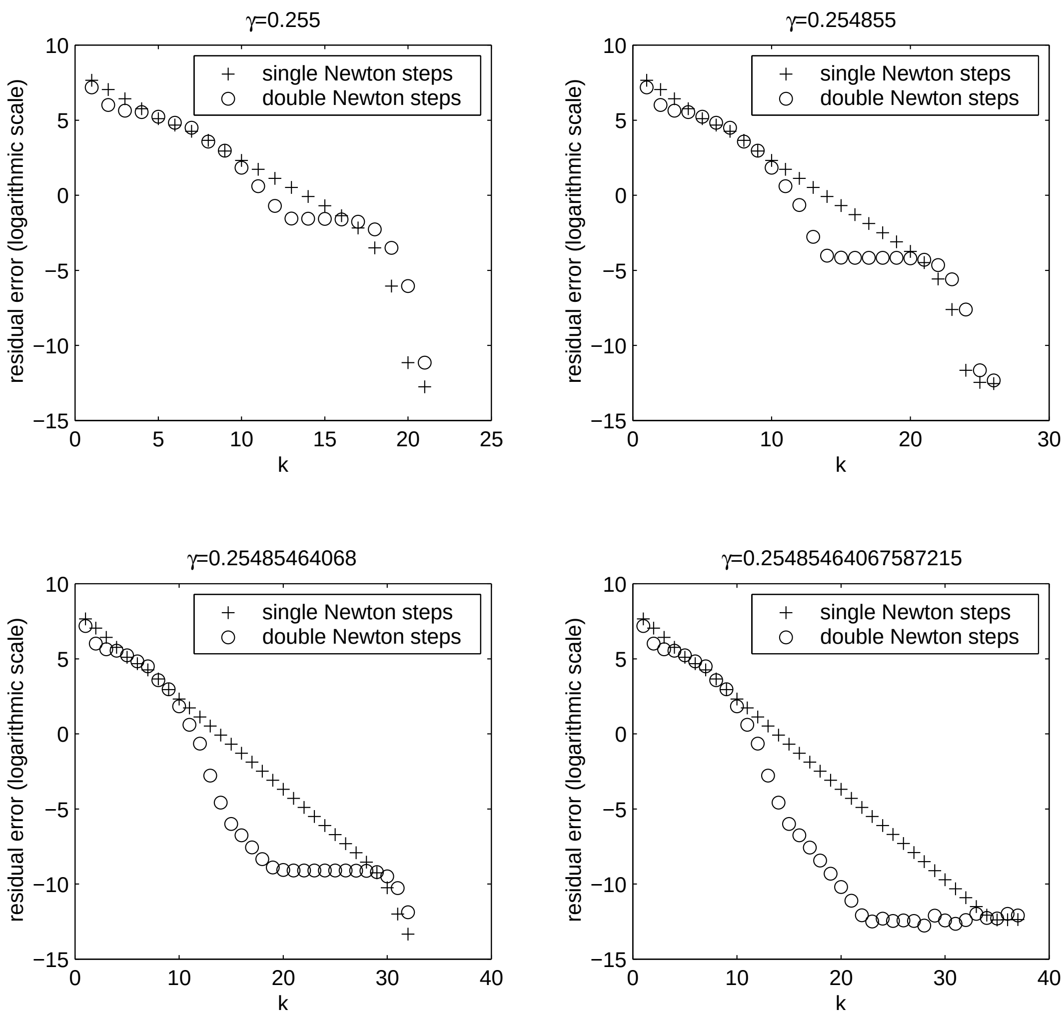
z. This effect is measured by the so-called *perturbation operator*, which maps the input signal $v \in L_2(\mathbb{R}_+)$ to the output signal $z \in L_2(\mathbb{R}_+)$. It has been shown in [9, 2] that the norm of this operator is less then γ for some given $\gamma > 0$, if and only if there exists a $Y > 0$ satisfying

$$\gamma^2 I + \sigma^2 B_{01}^* Y^{-1} B_{01} \quad > \quad 0$$

and

$$\begin{aligned} \mathcal{G}^\gamma(Y) \quad &= \quad -YA^* - AY - \sigma^2 Y A_0^* Y^{-1} A_0 Y - Y\tilde{C}^*\tilde{C}Y - B_2 B_2^* \\ &\quad\quad + B_1(\gamma^2 I + \sigma^2 B_{01}^* Y^{-1} B_{01})^{-1} B_1^* \\ &= \quad 0\,. \end{aligned}$$

The smallest γ for this constrained equation to be solvable is approximately given by $\gamma = 0.25485464067587215$. Again we visualize the residual errors $\|\mathcal{G}^\gamma(Y_k)\|$ and $\|\mathcal{G}^\gamma(Z_k)\|$ for the Newton iterates Y_k produced by single and Z_k produced by double Newton steps, when γ is close to this threshold value.



In all four figures we observe the effect that at some stage of the iteration the double Newton steps yield a better residual error than the single steps. But only, when γ is as close to the optimum as in the situation of the last figure, the error produced by an early double step is small enough such that the iteration can stop immediately.

References

[1] J. Ackermann. *Robust Control*. Springer-Verlag, 1993.

[2] T. Damm. State-feedback H^∞-type control of linear systems with time-varying parameter uncertainty. *Linear Algebra Appl.*, 351–352:185–210, 2002.

[3] T. Damm. *Rational Matrix Equations in Stochastic Control*. PhD thesis, University of Bremen, 2002.

[4] T. Damm and D. Hinrichsen. Newton's method for concave operators with resolvent positive derivatives in ordered Banach spaces. *Linear Algebra Appl.*, 363:43–64, 2003.

[5] R. Griesbaum. *Zur Stabilität dynamischer Systeme mit stochastischer Parametererregung*. PhD thesis, Fakultät für Maschinenbau, Universität Karlsruhe, 1999.

[6] C.-H. Guo. Newton's method for discrete algebraic Riccati equations when the closed-loop matrix has eigenvalues on the unit circle. *SIAM J. Matrix. Anal. Appl.*, 20(2):279–294, 1998.

[7] C.-H. Guo. Iterative solution of a matrix Riccati equation arising in stochastic control. Preprint, Department of Mathematics and Statistics, University of Regina, 2000.

[8] C.-H. Guo and P. Lancaster. Analysis and modification of Newton's method for algebraic Riccati equations. *Math. Comp.*, 67(223):1089–1105, 1998.

[9] D. Hinrichsen and A. J. Pritchard. Stochastic H_∞. *SIAM J. Control Optim.*, 36(5):1504–1538, 1998.

[10] P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Oxford, 1995.

[11] V. A. Ugrinovskii. Robust H^∞-control in the presence of stochastic uncertainty. *Internat. J. Control*, 71(2):219–237, 1998.

[12] W. M. Wonham. On a matrix Riccati equation of stochastic control. *SIAM J. Control*, 6:681–697, 1968.