Stochastic Approximation, Momentum, and Nash Play

H. Berglann[*] and S. D. Flåm[†]

April 3. 2002

Abstract. Main objects here are normal-form games, featuring uncertainty and noncooperative players who entertain local visions, form local approximations, and hesitate in making large, swift adjustments. For the purpose of reaching Nash equilibrium, or learning such play, we advocate and illustrate an algorithm that combines stochastic gradient projection with the heavy-ball method. What emerges is a coupled, constrained, second-order stochastic process. Some friction feeds into and stabilizes myopic approximations. Convergence to Nash play obtains under seemingly weak and natural conditions, an important one being that accumulated marginal payoffs remains bounded above.

Key words: Noncooperative games, Nash equilibrium, stochastic programming and approximation, the heavy ball method.

1. Introduction

Game theory has become an enormously important field of study [18], [22]. It now offers unity or coherence to many lines of inquiry. To wit, various social sciences, while steadily growing more game-theoretic, increasingly reckon Nash equilibrium as a focal point and key concept. Additional bonus and impetus comes with making those sciences more experimental [16].

The Nash solution concept nicely formalizes stable interaction among noncooperative players. In doing so it tends, however, to make larger demands on the players' knowledge and rationality than can easily be justified. Usually, to become a clever strategist one needs substantial learning or much experience. Therefore, Nash equilibrium begs some justification in dynamic terms [15], [19], [21].

Four features of human behavior then seem important. First, individuals always try to improve own welfare (or payoff). Second, they typically do not quite know all strategic possibilities, intentions or consequences. Third, they are likely to form local perspectives and approximations. Fourth, they hesitate in making quick and large adjustments.

[*] University of Bergen e-mail: helge.berglann@econ.uib.no Support from the Research Council of Norway is gratefully acknowledged.

[†] Corresponding author, University of Bergen and Norwegian School of Economics and Business Administration; e-mail: sjur.flaam@econ.uib.no. Thanks for support are due Ruhrgas, Røwdes fond and Meltzers høyskolefond.

To embody all these features in dynamics, and to accommodate exogenous uncertainty as well, we advocate and illustrate here use of stochastic approximation [6]. That vehicle, which leans heavily on differential equations, subscribes to a tradition that goes back to classical mechanics - and to Newton's claim that the initial state of a mechanical system determines its development.

The purpose of this paper is to put that claim to use for the study of stochastic games. Alternatively, one may read this paper as dealing with stochastic programming, parallel computation, or global optimization. Subsequent arguments are organized around a noncooperative stage game repeated time and again. Technicalities and proofs are found in the references.

2. The Game

There is a fixed, finite set I of players. Agent $i \in I$ is constrained to choose his strategy x_i from a nonempty compact convex subset X_i of a Euclidean space. He always seeks to improve own expected payoff $\pi_i(x_i, x_{-i}) := E\,\Pi_i(x_i, x_{-i}, \omega)$. Here $x_{-i} =: (x_j)_{j \neq i}$ denotes the part of the overall strategy profile $x = (x_i)$ that is controlled by i's rivals. The elementary event ω belongs to a complete probability space (Ω, σ, μ), with respect to which one takes the mathematical expectation E. Each bivariate function $(x, \omega) \mapsto \Pi_i(x, \omega) \in \mathbb{R}$ is concave, differentiable in x_i, and integrable in ω.

Of prime interest are points $x \in X := \Pi_{i \in I} X_i$ where each marginal payoff $m_i(x) := \frac{\partial}{\partial x_i}\pi_i(x)$ is normal to X_i - or quite simply, nil. That is, letting P_i denote the orthogonal projection onto X_i we seek a fixed point $x = (x_i)$ of the system

$$x_i \leftarrow P_i[x_i + s m_i(x)] \text{ for all } i \text{ and arbitrary } s > 0. \tag{1}$$

Any such fixed point is a Nash equilibrium. (1) amounts to a decentralized projected gradient procedure. It portrays fairly myopic parties, each trying to improve his linear approximation of own expected payoff. Modern, stochastic versions of such methods mirror two common aspects of human behavior: first, mean values (i.e. mathematical expectations E) are costly - and sometimes impossible - to compute; second, information concerning levels and gradients is readily available only at the current point. So, in our optic, letting $M_i(x, \omega) := \frac{\partial}{\partial x_i}\Pi_i(x, \omega)$ denote i's realized marginal payoff, one might hope to have almost sure convergence of the following stochastic process: For each i recursively posit

$$x_i^{k+1} := P_i\left[x_i^k + s_k M_i(x^k, \omega^k)\right]. \tag{2}$$

Here, at stage $k = 0, 1, \ldots,$ arrives a new event $\omega^k \in \Omega$, independently sampled according to the prescribed measure μ. Input at that stage k is also a positive stepsize s_k, selected a priori subject to

$$\sum s_k = +\infty \text{ and } \sum s_k^2 < +\infty. \tag{3}$$

The hope that process (2) converges is well founded provided $x \mapsto m(x) := [m_i(x)]_{i \in I}$ be globally monotone; see [9], [10], [11], [12], [13]. Otherwise, there are good reasons to be worried about convergence. Reflecting such worries our object here is to expand on gradient methods while preserving their many appealing properties.

Like (2) the procedure considered below does not presume much of foresight, experience, competence or optimization. In essence, it reflects iterated, noncoordinated pursuit of better payoffs. It does, however, modify the first-order gradient dynamics (2) by adding a second-order heavy-ball momentum, just like the harmonic oscillator of classical mechanics. Essentially, instead of assuming that player i pursues the gradient method $0 = m_i(x) - \dot{x}_i$ we posit that he rather drives the second-order process $\ddot{x}_i = m_i(x) - \dot{x}_i$. The latter must be suitably modified, of course, to account for discrete time, uncertainty, and constraints. This is done next.

3. Repeated Play

Let $\{\omega^k\}$ be a sequence of independent realizations of ω, each having distribution μ. As model of repeated play we advocate that iteratively at stages $k = 0, 1, \ldots,$ each individual i updates his current strategy x_i^k and velocity v_i^k by the rule

$$\left.\begin{array}{lll} x_i^{k+1} & := & P_i\left[x_i^k + s_k v_i^k\right] \\ v_i^{k+1} & := & v_i^k + P_i\left[x_i^k + s_k M_i(x^k, \omega^k)\right] - P_i\left[x_i^k + s_k v_i^k\right] \end{array}\right\} \tag{4}$$

As earlier, P_i denotes orthogonal projection onto X_i. Also like above, the parameter $s_k > 0$ is the stepsize used at stage k, selected a priori subject to (3). The initial points (x_i^0, v_i^0), $i \in I$, are determined by accident or historical factors better discussed in each particular setting.

To appreciate process (4) it helps to endow it with a clock that shows accumulated "time" $t_k := s_0 + \cdots + s_{k-1}$, $(\tau_0 := 0)$ at the on-set of stage k. Then, upon writing $x_i(t_k) := x_i^k$ and $v_i(t_k) := v_i^k$ we see that (4) assumes the form

$$\begin{array}{lll} \{x_i(t_{k+1}) - x_i(t_k)\}/s_k & := & \{P_i\left[x_i(t_k) + s_k v_i(t_k)\right] - x_i(t_k)\}/s_k \\ \{v_i(t_{k+1}) - v_i(t_k)\}/s_k & := & \left\{P_i\left[x_i(t_k) + s_k M_i(x(t_k), \omega^k)\right] - P_i\left[x_i(t_k) + s_k v_i(t_k)\right]\right\}/s_k \end{array}$$

Thus, since $s_k = t_{k+1} - t_k \to 0^+$, it turns out that behind (4) lurks - in expectation and the limit - a differential system

$$\left.\begin{array}{lll} \dot{x}_i & = & P_{T_i x_i}[v_i] \\ \dot{v}_i & = & P_{T_i x_i}[m_i(x)] - P_{T_i x_i}[v_i] \end{array}\right\} \tag{5}$$

Orthogonal projection $P_{T_i x_i}$ is here done onto the tangent cone $T_i x_i := \mathrm{cl}\mathbb{R}_+(X_i - x_i)$ of X_i at x_i. By a solution to (5) we understand an absolutely continuous profile $0 \leq t \mapsto [x(t), v(t)] = [x_i(t), v_i(t)]_{i \in I}$ which satisfies (5) almost everywhere. We suppose that the potential energy

$$0 \leq t \mapsto \int_0^t \sum_{i \in I} P_{T_i x_i(\tau)}[m_i(x(\tau))] \cdot \dot{x}_i(\tau) d\tau \tag{6}$$

remains bounded above along solution trajectories of (5). Following the arguments in [14] one may prove the following

Theorem 1. (Convergence of repeated play) Suppose system (5) has unique solution trajectories. Then, under the hypotheses above and the assumption that Nash equilibria are isolated, any discrete-time trajectory (x^k, v^k) generated by (4) must be such that x^k converges to a Nash equilibrium. □

Examples and illustrations of process (4), when deterministic, are found in [14]. We remark that at any stage k player i might, quite reasonably, first update his velocity v_i^{k+1} as prescribed and thereafter set $x_i^{k+1} := P_i\left[x_i^k + s_k v_i^{k+1}\right]$. In subsequent simulations we observe that this practice speeds up convergence.

4. Time-Homogeneous Play

When telling our tale about repeated play, we find it difficult sometimes to argue in favor of a time-inhomogeneous system. So, what happens if s_k is constant? Clearly, fixing this parameter is risky - and notably with stiff systems. To address that issue assume, for simplicity, that there are no constraints - or alternatively, that these have already been incorporated by means of suitable penalties. Setting $d_i = d_i(x, v_i, \omega) := M_i(x, \omega) - v_i$, iteration (4) then comes in autonomous, more tractable form

$$\left.\begin{array}{rcl} x_i^{k+1} & := & x_i^k + s v_i^k \\ v_i^{k+1} & := & v_i^k + s d_i^k \end{array}\right\} \qquad (7)$$

In simulations of (7), to hedge against stiffness and facilitate convergence, we replace d_i^k with a weighted sum $a_i d_i^k + b_i d_i^{k-1}$, using thus

$$\left.\begin{array}{rcl} x_i^{k+1} & := & x_i^k + s v_i^k \\ v_i^{k+1} & := & v_i^k + s(a_i d_i^k + b_i d_i^{k-1}) \end{array}\right\} \qquad (8)$$

The parameters a_i, b_i could account for accumulated learning on how to adapt in a complex dynamic environment.[1] We remark that the second equation in (8) is strikingly similar to a control algorithm commonly used in process industry, namely: the so-called Proportional-Integral-Controller; see [5], [17], [23], [24]. We take advantage of this to find appropriate a_i, b_i-parameters in the example below:

5. An Example

Let each $i \in I$ here be a Cournot oligopolist [8] who supplies the quantity $x_i \geq 0$ of a homogeneous, perfectly divisible good to a common market. Thereby he receives sales revenues px_i and incurs (differentiable) production costs $c_i(x_i)$. The price is determined by a smooth inverse demand curve which is subject to stochastic

[1] In fact, if agent i were new to the kind of dynamic process in question, his optimal behavior might entail experimentation to determine the said parameters. Most likely agents would learn from each other and, if possible, imitate those who do well. Henceforth assume that each i has previous experience, perhaps from similar processes, and has appropriately tuned his a_i and b_i.

‡uctuations. Specifically, the realized price equals $p = !P(Q)$ with $E! = 1$ and $Q = \mathbf{P}_{i2I} x_i$. Thus

$$E ¦_i (x_i; x_{¡i}; !) = P(Q)x_i \; ¡ \; c_i(x_i) \text{ and } EM_i (x_i; x_{¡i}; !) = P(Q) + P^0(Q)x_i \; ¡ \; c_i^0(x_i)$$

The figures below depict individual supply x_i over stages k with constant stepsize $s = 1$, as generated by (8). There are ten players ($jIj = 10$), $!$ has a lognormal distribution with standard deviation $0{:}3$, $P(Q) = 10 \; ¡ \; Q$, $c_i(x_i) = x_i$, $x_i^0 = 1$, $v_i^0 = 0$ and finally, $d_i^0 = 0$ for all i. The resulting Nash Equilibrium x is unique with all $x_i = 0{:}818$.

The resemblance with controller algorithms made us look for methods to determine e¢cient values of a_i and b_i in the literature on Control Engineering. The wide spread use of such algorithms not withstanding, there exists no generally accepted method for how to tune these parameters. The empirical method developed by Ziegler and Nichols (1942) [24] is still holds good ground and has the great advantage of requiring very little information [23]. The said method gave the values of a_i and b_i that label Figure 1 and are used by all agents. The scantly dotted line shows behavior in the deterministic case (when $! \, , \, 1$) while in the more solid curve $!$ is sampled anew for each k.

Figure 2 brings out responses when all agents half the values a_i and b_i employed in Figure 1. Absent uncertainty, the time needed to reach a steady level now becomes longer. Present uncertainty, (using the same series $!^k$ as in Figure 1) it causes less ‡uctuations than before.

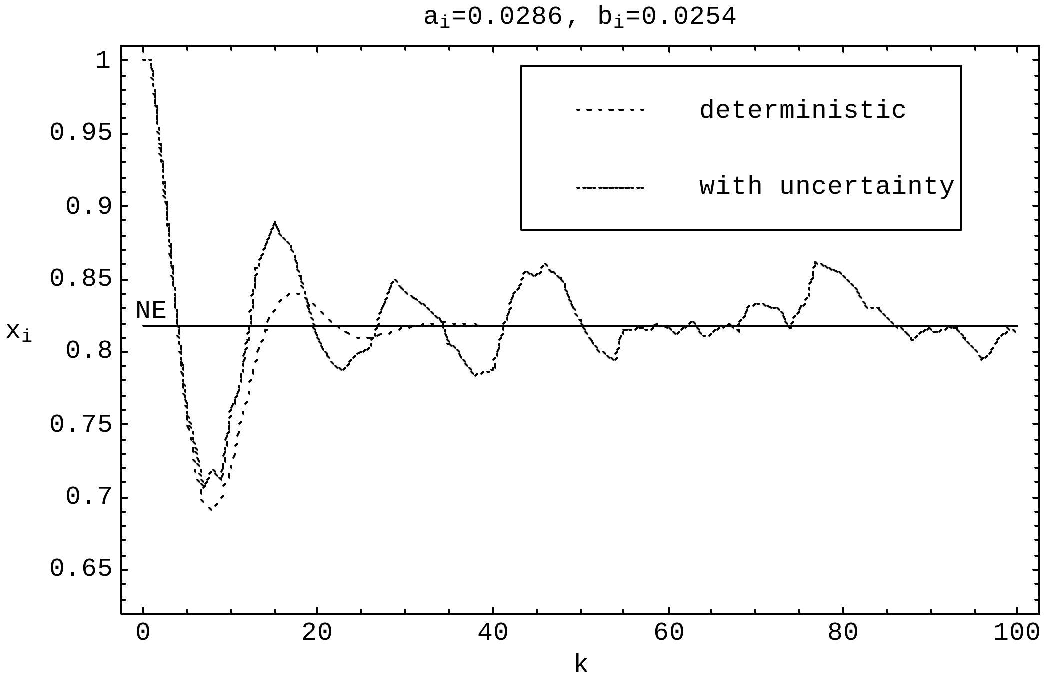

Figure 1: Supply x_i over stages k when all players employ parameters a_i and b_i determined by Ziegler-Nichols Method.

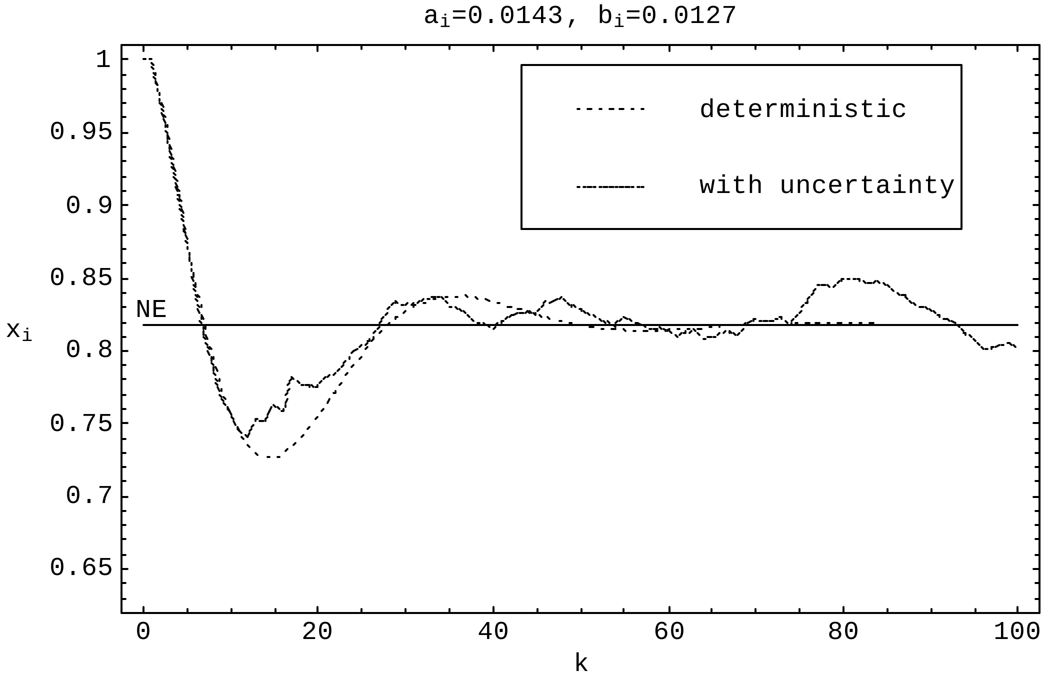

Figure 2: Supply x_i over stages k when all players employ parameters a_i and b_i with values half of the size determined by Ziegler-Nichols Method.

Figure 3 illustrates what happens when parameters $a_i; b_i$ differ across agents. The five first players use values listed in Figure 1; the others use those mentioned in Figure 2. Members of the first group adapt fastest initially. Differences across agents make for slower convergence in the deterministic case.

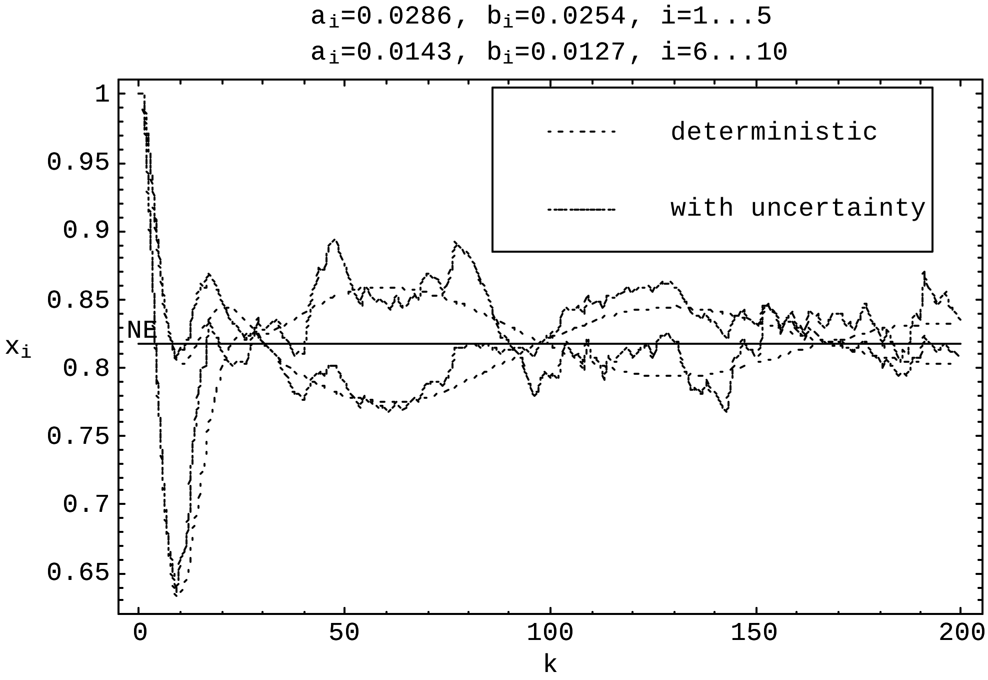

Figure 3: Supply x_i over stages k with different $a_i; b_i$. Group 1 uses parameters a_i and b_i determined by Ziegler-Nichols method - while group 2 half these values. Members of the first group adapt fastest initially.

6. Concluding Remarks

We have presented (4) as a tale about repeated play of noncooperative, constrained games. A main motivation behind the heavy-ball philosophy was to reach beyond instances where the solution set is connected. A supplementary aim - referred to as equilibrium selection [21] - amounts to exploit uncertainty or blurred data so as to arrive at particularly stable solutions. In fact, randomness - if not already a key ingredient - could artificially be introduced to escape from unstable equilibria.

Clearly, when I is a singleton, this paper fits the frames of single-agent optimization under uncertainty. In that regard (4) has something to offer in three respects. First, the heavy-ball method of Polyak [20] has, to our knowledge, not come fully into stochastic programming. Second, process (7) is amenable to parallel computing [7]. Third, following Attouch et al.[3], the same process is applicable for global optimization - or the selection of "good" stationary points; see also [1], [2], [4]. Approach (8) may require, for its efficient operation, some auto-tuning of the parameters $a_i; b_i$: Appropriate routines to that effect are found in the engineering literature on control; see for instance [5], [23].

References

[1] F. Alvarez, On the minimizing property of a second order dissipative system in Hilbert spaces, SIAM J. Control Opt. 38, 4, 1102-1119 (2000).

[2] F. Alvarez and J. M. Pérez, A dynamical system associated with Newton's method for parametric approximations of convex minimization problem, Appl. Math. Optm. 38, 193-217 (1998).

[3] H. Attouch, X. Goudou, and P. Redont, The heavy ball with friction method I, The continuous dynamical system: Global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system, Communications in Contemporary Mathematics 2, 1, 1-34 (2000).

[4] H. Attouch and P. Redont, The second-order in time continuous Newton method, in M. Lassonde, Approximation, Optimization and Mathematical Economic, Physica-Verlag, Heidelberg 4-36 (2001).

[5] R. Bandyopadhyay and D. Patranabis, A fuzzy logic based PI auto-tuner, ISA Transactions 37, 227-235 (1998).

[6] M. Benaim, A dynamical system approach to stochastic approximation, SIAM J. of Control and Optimization 34, 437-472 (1996).

[7] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, New York (1989).

[8] A. Cournot, Recherches sur les principes mathématiques de la théorie des richesses, Riviere & Cie, Paris (1838).

[9] S. D. Flåm, Approaches to economic equilibrium, Journal of Economic Dynamics and Control 20, 1505-1522 (1996).

[10] S. D. Flåm, Restricted attention, myopic play, and the learning of equilibrium, Annals of Operations Research 82, 473-482 (1998) .

[11] S. D. Flåm, Learning equilibrium play: a myopic approach, Computational Optimization and Appl.14, 87-102 (1999) .

[12] S. D. Flåm, Repeated play and Newton's method, Int. Game Theory Review 2, 141-154 (2001).

[13] S. D. Flåm, Approaching equilibrium in parallel, in D. Butnariu, Y. Censor and S. Reich (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, North-Holland 267-278 (2001).

[14] S. D. Flåm and J. Morgan, Newtonian mechanics and Nash Play, manuscript (2001).

[15] D. Fudenberg and D. K. Levine, The Theory of Learning in Games, MIT Press, Cambridge Mass. (1998).

[16] J. K. Goeree and C. A. Holt, Ten little treasures of game theory and ten intuitive contradictions, The American Economic Review 91, 5, 1402-1422 (2001).

[17] C. C. Hang, K. J. Åstrøm, Q. G. Wang, Relay feedback auto-tuning of process controllers – a tutorial review, Journal of Process Control 12, 143-162 (2002).

[18] J. Hofbauer and K. Sigmund, Evolutionary Games and Population Dynamics, Cambridge University Press (1998).

[19] J. Hofbauer, From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS, Selection 1, 81-88 (2000).

[20] B. T. Polyak, Some methods of speeding up the convergence of iteration methods, Z. VyCisl. Math. i Mat. Fiz. 4, 1-17 (1964).

[21] L. Samuelson, Evolutionary Games and Equilibrium Selection, The MIT Press, Cambridge, Massachusetts (1997).

[22] J. Watson, Strategy, Norton, New York (2002).

[23] K. J. Åström, H. Panagopoulos and T. Hägglund, Design of PI Controllers based on Non-Convex Optimization, Automatica 34, 5, 585-601 (1998).

[24] J. G. Ziegler and N. B. Nichols, Optimum settings for automatic controllers, Trans. ASME 64, 759-768 (1942).