Optimal behavior of consecutive chemical reactions
A⇔B⇔C

Thor A. Bak[1], Peter Salamon[2], Bjarne Andresen[2,3]

[1]Chemistry Laboratory III, University of Copenhagen, Universitetsparken 5, DK 2100 Copenhagen Ø, Denmark

[2]Department of Mathematical Sciences, San Diego State University, San Diego, CA 92182 7720, USA

[3]Ørsted Laboratory, University of Copenhagen, Universitetsparken 5, DK 2100 Copenhagen Ø, Denmark

Abstract

The consecutive chemical reactions A⇔B⇔C are analyzed both numerically and analytically using temperature as the control variable. Starting with pure A and maximizing the yield of B at the end of the given process duration, all optimal paths start with a branch at infinite temperature. A curve on which switching from this temperature to lower temperatures is possible is derived. The production rate of B at the end of the reaction is found to equal the incremental gain of B by increasing the duration when the path is re-optimized for this new duration. For given parameters there is a unique "maximum useful time" which results in the largest yield of B possible. This duration may be infinite. If a duration longer than this is specified, all reactions should be shut off for that excess amount of time, a situation which makes most optimization routines become unstable.

1. Introduction

Already half a century ago, several authors [1-11] considered the question of improving the yield of a chemical reaction taking place in a tubular reactor by adjusting the temperature in various parts of the reactor. The papers most pertinent to the present work are those of Bilous and Amundson [3] and Aris [8-11].

By considering a reaction in a tubular reactor through which the reacting material is pumped and along which the temperature varies, we are in effect considering a reaction at a temperature which varies with time. If there is

only one chemical reaction going on, the more or less obvious answer to the problem of getting maximum yield is to adjust the temperature so that the reaction rate is as large as possible at each point of the reactor. This in most cases just means keeping the temperature as high as possible. In the following, we let the temperature be infinite, but this just means "as high as possible, other conditions taken into account".

If there are two competing reactions A→B and A→C for which B is the desired product and C is a waste product, the situation is not so simple, since by increasing the rate constants for the first reaction we automatically also increase the rate constants of the second thereby letting A go directly to waste. If there are two consecutive reactions A→B→C for which as before B is the desired product and C the waste, the situation is additionally complicated and furthermore so if the reverse reactions are also taken into account.

The latter system is a crude model for the operation of a tunnel oven where B is the desired product and C is deadburned material. Thus for gypsum A is $CaSO_4·2$ H_2O, B is burned gypsum $CaSO_4·1/2$ H_2O which readily reabsorbs water under solidification, while C is deadburned gypsum $CaSO_4$ which only very slowly reacts with water. Another example is the production of cement. The actual reactions going on in an industrial oven are quite complicated and we will not attempt to reproduce them. Our model only describes the general features using relative values of rate constants and activation energies to exemplify the phenomenon.

Aris [8] solved the problem of two consecutive reactions using the method of dynamic programming rather than functional differentiation used by Bilous and Amundson [3]. He thereby also resolved two questions left open by them. In this paper we use the method of optimal control theory, which in the meantime has become available, on the problem A⇔B⇔C. Our analysis confirms the result of Amundson and Aris that one should not merely adjust the temperature to its highest value throughout the reaction. At a certain stage, when the back reaction begins to be important, the temperature must be lowered. We analyze this feature in detail. The method we use is more powerful than those used previously and it may be applied to other and more complicated reaction schemes. However, except when further simplifications are introduced (as shown in Section 5), an analytical solution cannot be obtained.

2. System

We consider the system of first order chemical reactions

$$A \underset{k_2}{\overset{k_1}{\Leftrightarrow}} B \underset{k_4}{\overset{k_3}{\Leftrightarrow}} C \qquad (1)$$

with rate constants k_i, $i=1...4$ depending on temperature T through Arrhenius expressions,

$$k_i = s_i \, e^{-E_i / kT} \qquad (2)$$

with activation energies E_i and collision factors s_i. These are taken to be constants whereas the temperature T varies with time because, as mentioned in the introduction, the reacting mixture is pumped through a tube along which the temperature varies. From now on we forget the experimental setup and just assume we have a reaction mixture for which the temperature can vary with time and take this to be our control variable.

Starting from pure A, i.e. [A]=1, [B]=[C]=0, our objective is to produce as much B as possible within a given duration τ by varying the temperature appropriately over the time span $[0,\tau]$. We will not put any restrictions on the temperature except that it must be positive, and we assume that it can be changed freely and instantaneously as needed.

3. Optimization procedure

We use optimal control theory [12] to derive the optimal temperature path. In order to avoid the omnipresent exponential functions in the rate constants, it will prove convenient to change the control variable from T to

$$u = e^{-1/kT} \qquad (3)$$

with u restricted to the interval $[0,1]$ corresponding to the temperature interval $[0,\infty]$ so that

$$k_i = s_i \, u^{E_i} \qquad (4)$$

Our objective function to be maximized is thus

$$B(\tau) = \int_0^\tau \frac{dB}{dt} \, dt \qquad (5)$$

subject to the three constraints

$$\frac{dA}{dt} = -k_1 A + k_2 B \qquad (6)$$

$$\frac{dB}{dt} = k_1 A - k_2 B - k_3 B + k_4 C \stackrel{def}{\equiv} F \tag{7}$$

$$A + B + C = 1 \tag{8}$$

and with initial conditions

$$A(0) = 1, \quad B(0) = C(0) = 0 \tag{9}$$

Here and in the following we have skipped the conventional square brackets when writing concentrations, i.e. we use e.g. A instead of [A]. The third rate equation dC/dt is a consequence of dA/dt, dB/dt and mass conservation, eq. (8) and thus automatically satisfied.

We will at once reduce the set of equations further by substituting C from eq. (8) into eq. (7) so that the optimization only has two constraints, eq. (6) and

$$\frac{dB}{dt} = (k_1 - k_4)A - (k_2 + k_3 + k_4)B + k_4 \tag{10}$$

The Hamiltonian for this system is

$$H = \mu_0[(k_1 - k_4)A - (k_2 + k_3 + k_4)B + k_4] + \lambda(-k_1 A + k_2 B)$$
$$+ \mu_1[(k_1 - k_4)A - (k_2 + k_3 + k_4)B + k_4] \tag{11}$$

where the Lagrange multiplier μ_0 associated with the objective function must be constant and can only have the values 0 or 1. However, since the final values of A and B are free so that $\lambda(\tau) = \mu_1(\tau) = 0$, $\mu_0 = 0$ is not a possibility as it would make the final value of the total conjugate vector equal to zero. Upon combining the two Lagrange multipliers μ_0 and μ_1 into a single $\mu = \mu_0 + \mu_1$, the Hamiltonian becomes

$$H = \lambda(-k_1 A + k_2 B) + \mu[(k_1 - k_4)A - (k_2 + k_3 + k_4)B + k_4] \tag{12}$$

The Hamiltonian equations to be solved are then

$$\frac{\partial H}{\partial A} = -\frac{d\lambda}{dt} \qquad \frac{\partial H}{\partial B} = -\frac{d\mu}{dt}$$

$$\frac{\partial H}{\partial \lambda} = \frac{dA}{dt} \qquad \frac{\partial H}{\partial \mu} = \frac{dB}{dt} \tag{13}$$

under the boundary conditions eq. (9) and

$$\lambda(\tau) = 0 \qquad \mu(\tau) = 1 \tag{14}$$

Thus we have initial conditions on the state variables but final conditions on the conjugate variables, corresponding to the final states being free to vary to achieve the optimality. The two lower equations in (13) which specify the time evolution of the state variables A and B are already stated above as eqs. (6) and (10). Finally, the control $u(t)$ is determined so as to maximize H:

$$u = \arg\max_u H \quad, \tag{15}$$

where argmax H means the value of the argument u of H which results in the maximum possible value of H. The Hamiltonian H depends on the "temperature" u through the rate constants k_i.

In general the Hamiltonian equals the incremental yield of the objective function by increasing the duration of the process,

$$H = \frac{dB^*}{d\tau}, \tag{16}$$

where B^* is the optimal value of the objective function B at the final time τ. At this particular point $dB^*/d\tau$ and dB/dt are the same reaction rate. This follows by substituting the boundary conditions eq. (14) into the Hamiltonian expression eq. (12) and identifying the resulting expression as dB/dt, eq. (10). Thus we can follow the approach to the maximum possible product $B^*(\tau)$ as τ is increased either by following the value of the Hamiltonian or of dB/dt at the end of the process. This observation is true generally, provided the objective function coincides with the final value of one of the state variables, and none of the other state variables have constrained final values, a common occurrence for chemical systems. For unconstrained duration τ the Hamiltonian vanishes and so will the final reaction rate.

We can glean further understanding from the unconstrained duration case. If there exists a finite optimal duration, say τ^*, then any time greater than τ^* can do equally well. This follows since for any time $\tau > \tau^*$ we can augment the process by setting the control $u=0$ for time τ-τ^*. Note that u=0 implies that the Hamiltonian is zero and all of the dynamical equations give zero derivatives – as expected on physical grounds, the process is frozen. Such u=0 branches can be added to any optimal process and still have the results satisfy all the conditions of optimality. In fact for solutions corresponding to H=0, such branches can be spliced in anywhere along the process. Once the constrained time τ exceeds the minimum optimal time, τ^*_{min}. This fact serves to confound the numerical algorithms for the solution as described further below.

4. Numerical optimization

For the numerical examples of this boundary value problem we solve eqs. (13) using Krotov's method [13] in which alternately the state differential equations (in *A* and *B*) are solved in the forward time direction and the conjugate differential equations (in λ and μ) are solved in the backward direction. In this way each set of equations starts with its given boundary conditions. After each iteration back and forth the *u* which maximizes *H* is determined at each point along the trajectory. This method is much more efficient than the conventional "shooting method" and usually converges in just a few iterations.

The optimization converges quickly as long as τ is not too long, i.e. as long as time is at a premium so that it must be used carefully. For long reaction times B will react further to become C, depleting the desired product. This can be prevented only by shutting off all reactions by reducing the temperature to zero. However, these periods of standstill can equally well be placed anywhere during the process, either as one large period or as several shorter ones, leaving the optimization in a dilemma as to what to do. Mathematically this shows up as unstable equations which cannot converge but keep switching between zero and non-zero temperature. In these situations the Hamiltonian has multiple minima for *u* in the physical interval [0,1], and small variations during the iterative solution may easily flip the control between very different values, causing erratic behavior. We were only able to obtain numerical solutions for duration $\tau < \tau^*_{min}$.

Figure 1ab show the optimal time sequences of the concentrations *A*, *B*, and *C* as well as the conjugate variables λ and μ for one such calculation. Frame (c) shows the time-independent plot of *A* vs. *B*. Frame (d) contains the control *u* ("temperature") which produces the optimum. The activation energies and collision factors used are $E=(1, 2, 2, 1)$ and $s=(0.1, 0.1, 0.1, 0.1)$, corresponding to B being the most stable component and the equations integrated for a fairly long time. This is a textbook example with infinite initial temperature, gradually being reduced so as to push the equilibrium toward the desired, energetically favored product. This is an example with infinite maximum usable time τ^*_{min} as is the case for any example[1] in which $k_1 = k_4$.

---

[1] In this case it follows that the adjoint variable λ must vanish identically and thus the $H=0$ case forces $dB/dt \equiv 0$ contradicting optimality in finite time.

Figure 2 uses E=(1, 2, 1, 3) and s=(0.7, 0.7, 0.7, 0.7). This represents a system where B is more stable than A and C still more stable. Thus without control the final product would be C. The optimal path calls for a relatively long period at infinite temperature after which the temperature jumps to a much lower value to prevent loss into C (see frame d).

Figure 3 with E=(2, 1, 3, 1) and s=(0.6, 0.6, 0.6, 0.6) is the opposite case where A is the most stable substance. Again the optimal path starts with a fairly long period at infinite temperature but then gradually breaks away to lower values as B should be favored over C (frame d). In this case for which the optimal temperature moves gradually off down from infinity, we can derive a switching curve shown in frame (c). The cooling begins when the phase trajectory of A vs. B crosses this curve.

A couple of interesting general features should be noted for later discussion.

1. All optimal solutions start with a period of u=1 (T infinite) after which the curve suddenly breaks off to lower temperatures when back reaction begins to be important. This feature is proved in general in the next section.

2. The final concentrations of A and C are not the lowest / highest seen throughout the calculation. Rather, the reactions seem to have gone a little too far and then return partly to A and B, respectively. This may appear counterproductive, but as long as more B is produced from C than returns to A, the balance for B is still positive.

5. Analytical optimization

Analytical solution of eqs. (13) can be facilitated by replacing time as the independent coordinate with another quantity which increases (or decreases) monotonically during the process. This is a standard technique for reducing the dimensionality of the problem and often yields to analytic attack in the lower dimensional version. For chemical reactions, one would often use the extent of one of the reactions in eq. (1) as the new time variable. However our last numerical observation above, that the reactions often reverse briefly in A or C toward the end of the allowed period τ invalidates this choice since the parameter is no longer monotonic in time. Instead we can use the concentration of the desired product B itself. To our knowledge this substitution has not been used before. It effectively reverses the roles of t and B by defining B as our independent variable and τ as our objective function. This switching of a constraint and the objective function is one of the standard dual problems in optimization theory [14] and is routinely used in thermodynamics, e.g. when alternatively maximizing the entropy of a system

while keeping the energy constant or minimizing the energy while keeping the entropy constant. Whereas originally we sought the maximum final B for given duration τ, we now correspondingly search for the minimum duration

$$\tau = \int_0^{\overline{B}} \left(\frac{dB}{dt}\right)^{-1} dB = \int_0^{\overline{B}} \frac{dB}{F} \tag{17}$$

for a fixed given production $\overline{B}$ of B. The original question is the inversion of this relationship, i.e. $B(\tau)$ rather than $\tau(\overline{B})$.

We are now left with only one constraint,

$$\frac{dA}{dB} = \frac{-k_1 A + k_2 B}{F} \tag{18}$$

giving a simpler Hamiltonian,

$$H = -v_0 \frac{1}{F} + v\frac{-k_1 A + k_2 B}{F} \tag{19}$$

and only one pair of Hamilton's equations, eq. (18) and

$$\frac{dv}{dB} = -\frac{\partial H}{\partial A} = -\frac{v[(k_1 k_3 + k_1 k_4 + k_2 k_4)B - k_1 k_4] + (k_1 - k_4)}{F^2}. \tag{20}$$

 Alternatively, the conjugate variable $v(B)$ can be obtained from the constancy of the Hamiltonian. Like before we must solve the coupled equations (18) and (20) and obtain the "temperature" u as argmax H.

Solving eqs. (18) and (20) is still not simple so we will find the boundary branch $u=1$ ($T=\infty$) and prove that all optimal solutions will start on such a branch. Differentiating H with respect to u initially where $A=1$ and $B=0$, we find

$$\frac{dH}{du} = v_0 \frac{E_1 u^{-E_1 - 1}}{s_1} \tag{21}$$

which is positive in [0,1], so we can conclude that $u=1$ is the only maximum of the Hamiltonian as long as v_0 is not in the constraint-dominated case of zero, i.e. as long as "time" has effect on the yield of the reaction. This means that all (non-constraint-dominated) solutions will begin with a segment of $u=1$.

Fot this initial branch, we need to solve the constraint equation (18) which with $u=1$ reads

$$\frac{dA}{dB} = \frac{-s_1 A(B) + s_2 B}{(s_1 - s_4)A(B) - (s_2 + s_3 + s_4)B + s_4}. \tag{22}$$

With the *A*-term in the denominator we cannot solve this analytically, but if we assume that $s_1 = s_4$, the equation becomes linear and we find

$$A(B) = \left[s_3 \left(\frac{s_4 - (s_2 + s_3 + s_4)B}{s_4} \right)^{\frac{s_4}{s_2 + s_3 + s_4}} + s_2(1 - B) \right] \Bigg/ (s_2 + s_3). \tag{23}$$

How long should the system remain at $u=1$? This is a difficult but chemically interesting question. The $u=1$ trajectory is followed as long as this value of u maximizes the Hamiltonian. This situation can change either abruptly by a better maximum showing up at another value of u resulting in a jump to that temperature or gradually by the minimum moving off from the boundary $u=1$. The first of these cases is illustrated by the example in Fig. 2; the second by the examples of Figs. 1 and 3. In the case where the maximum moves gradually off the $u=1$ solution, we can find the switching curve, i.e. a relationship between the A and B concentrations that is necessary for this to happen provided the maximum usable time τ^*_{min} is finite. We find the point at which the process should leave the boundary by substituting $u=1$ into $H = dH/du = 0$ and solving the resulting system of equations. The finiteness of τ^*_{min} is used in assuming that $H=0$. The resulting switching curve is shown in Fig. 3 which is the only one of our 3 examples for which the derivation applies. In these equations *B* acts as "time", which is fortuitous for the phaseplot of $A(B)$ in Fig.3.

6. Conclusions

The consecutive chemical reactions A⇔B⇔C have been analyzed both numerically and analytically. Starting with pure A and maximizing the amount of B at the end of the given process duration, we find that all optimal paths start with a segment at infinite temperature ($u=1$). Depending on the energy barriers and the collision factors, the optimal temperature may subsequently switch to either a finite temperature (internal optimum) or zero temperature ($u=0$) if the "maximum useful time" is exceeded. This is a standard turnpike structure. The time rate of change of *B* is positive throughout the process. For certain sets of parameter values the production rate *dB/dt* will approach zero asymptotically for long process durations, for others it will become zero for finite durations. If such an internal zero exists, the optimization becomes indeterminate for longer durations since the required periods of shutting off the reaction can equally well appear anywhere during the process.

The present line of development should be considered part of the continuing quest to answer R. Stephen Berry's original vision of finite time thermodynamics [15] as the exploration bounds on what can be achieved in finite time. While the problem discussed in this article does not concern the energetics of the process directly, it is an attempt to find how much of A can be turned into B in a given time and thus belongs to the realm of finite time thermodynamics. While we have found several interesting conclusions regarding this simple system, much more remains unanswered.

Acknowledgements

References

[1]   D. Annable, Chem. Eng. Sci. **1**, 145 (1952).

[2]   N. R. Amundson, De Ingenieur No 37, 1 (1955).

[3]   O. Bilous, N. R. Amundson, Chem. Eng. Sci. **5**, 81 115 (1956).

[4]   K. G. Denbigh, Chem. Eng. Sci. **8**, 125 132 (1958).

[5]   F. Horn, U. Troltenier, Chemie Ing. Techn. Nr. 6, 382 393 (1960).

[6]   F. Horn, Z. Elektrochemie **65**, 209 222 (1961).

[7]   F. Horn, Chem. Eng. Sci. **14**, 77 89 (1961).

[8]   R. Aris, Chem. Eng. Sci. **13**, 18 29 (1960).

[9]   R. Aris, Chem. Eng. Sci. **13**, 197 206 (1961).

[10] R. Aris, Z. Elektochemie **65**, 229 244 (1961).

[11] R. Aris, R. S. H. Mah, Chem. Eng. Sci. **19**, 541 553 (1964).

[12] G. Leitmann, The Calculus of Variations and Optimal Control, Plenum Press, New York, 1981.

[13] Krotov V.F. Computing methods for solution and optimization of the controlled systems of equations I, II. Izvestia of USSR Acad. Nauk, Ser. Techn. Cybern. N5, N6, 1975.

[14] A.L. Peressini, F.E. Sullivan, and J.J. Uhl, Jr. The Mathematics of Nonlinear Programming. Springer Verlag, New York, 1988, p. 63.

[15] "Thermodynamics in Finite Time", B. Andresen, P. Salamon, and R.S. Berry, Physics Today, September 1984. (or should we refer to his recent book with Anatoly, Stan, and co.?)

Figure captions

Figure 1. Optimal time sequences of the concentrations A, B, and C (frame a: dashed, solid, dotted) as well as the conjugate variables λ and μ (frame b: dashed and solid) for activation energies $E=(1, 2, 2, 1)$ and collision factors $s=(0.1, 0.1, 0.1, 0.1)$. Frame c shows the phase relationship, A as function of B, for the optimal path Frame d shows the time evolution of the control u ("temperature") which produces the optimum yield.

Figure 2. Optimal solution for activation energies $E=(1, 2, 1, 3)$ and collision factors $s=(0.7, 0.7, 0.7, 0.7)$. The nomenclature is the same as in Fig. 1.

Figure 3. Optimal solution for activation energies $E=(2, 1, 3, 1)$ and collision factors $s=(0.6, 0.6, 0.6, 0.6)$. The nomenclature is the same as in Fig. 1.