



SURVEY OF OPTIMAL CONTROL AND MODEL PREDICTIVE CONTROL WITH STATE ESTIMATION AND A REAL TIME APPLICATION

László DÁVID,¹ Katalin GYÖRGY,² László-Alpár GALACZI³

Sapientia Hungarian University of Transylvania, Faculty of Technical and Human Sciences Târgu Mureş, Department of Electrical Engineering, Târgu Mureş, Romania

¹ ldavid@ms.sapientia.ro,

² kgyorgy@ms.sapientia.ro

³ galaczil@yahoo.com

Abstract

The optimal control and its limited version namely the model predictive control represent one of the most important nonlinear control alternatives nowadays. The success of them are also proven in many practical applications. These can provide for several industrial applications the optimal trajectory calculation as well as calculation of the real-time control signal. One successful version of this is Generalized Predictive Control (GPC). A big advantage of these control algorithms is that they solutions are able to take into account the limitations of the inputs, and the states. In some cases, it is important to know the mathematical model chosen and the complete state information. Otherwise, the model can be estimated during the operation. Our study shows through the control of the cathode heating of a high-power electron beam device the self-tuning adaptive control thus constructed. Using a suitable dynamic model and an extended Kalman estimator, we determine the estimated temperature of the two cathodes during operation and the saturation electron current, which ensures the maximum cathode life. The practical application was tested on a CTW 5/60 type electron gun.

Keywords: optimal control, Kalman filter, model predictive control, optimal trajectory.

1. Introduction

It is a known fact that today, due to the spread of digital embedded systems (although the numerical control algorithms are very often used), a significant proportion of these operate with a classical PID or equivalent control algorithm. It can also be observed that the so-called adaptive PID algorithms have become very popular because they are easy to implement and very simple to review. However, for high standard controls, the maximum one-step prediction provided by PID is not always adequate in practice. In this case the so-called minimum variance and model-based predictive controls can be useful, where the horizon can be chosen by the designer. In these contexts, the equivalent of a PID control algorithm is the General Predictive Control (GPC). This is easily feasible in practice and - thanks to the prediction horizon and the fact that it can obtain an adaptive algorithm (supplemented with a parameter estimator) - it can certainly be a good alternative to the mentioned control algorithms. In practice, a GPC algorithm together with a least-squares parameter estimation algorithm already provides adaptive control. The parameter estimation block specifies the regression parameters corresponding to the current operating point, and based on this, the GPC control calculates the corresponding sequences of the control signal.

Naturally, the question here is whether we have an algorithm that can perform both tasks in a given sampling time, which is often of the order of milliseconds. Well, there are many processes where an adaptive GPC algorithm is unable to do this. Moreover, in the case of nonlinear systems (where the estimated parameters are valid only in a narrow environment of the operating point and therefore the estimation has to be performed in all samples), these control implementations can be more sensitive to the calculation times. The question is, what would be the algorithm that would be able to perform the estimation over the entire operating range, even for nonlinear systems, and thus would not have to waste expensive computational time during operation?

One such variant of the GPC control would be the Neural GPC control. This method performs the regression estimation with a neural network and not with one of the ARMAX, ARIMAX, CARI-MAX models. Naturally, this is also a regression estimate, known in the literature as the NARMAX estimate; with this, the corresponding GPC control signal can calculated based on a preliminary estimate. Although many have dealt with this (for example even Professor Norgaard who introduce the NARMAX, has developed a software package), yet a significant number of applications apply the method only to linear processes. This work aims to study the possible advantages and disadvantages of these control methods in the case of a strongly nonlinear process.

Our test process is a cathode-assembly heating controller, applied to high power electron-beam equipment. This was originally accomplished with a PID controller, and cannot be used especially at very high-power conditions. As a conclusion, it was possible to control up to a maximum of 80-90 W, in stable mode, which in many cases was not enough to achieve the maximum electron beam processing current.

A very important chapter in the development of control algorithms can be classified as the adaptive control algorithms, such as the self-tuning control algorithm. Aström and Wittenmark [1], originally proposed only for SISO systems as a Minimal Variance Control (MVC) algorithm supplemented with a parameter estimation part. For multivariable MIMO systems, Keviczky, Hetthéssy, Hilger, and Kolostori [2]. developed this method, among others. Clarke and Gawthrop, the developers of GPC control, also noted that in these cases, in addition to knowing the exact dead time, here is required a non-zero value of the control signal limit of the objective function. Moreover, it has been observed that if in the design phase - the polynomial defining the

control signal of the used regression model $B(z^{-1})$, is neglected then the designed closed-loop control will be unstable if this polynomial has unstable roots. They proposed the introduction of an auxiliary output that modifies the structure of the used regression model [3]. Unfortunately, this method did not work either. The one-step estimation used here does not consider the dead time as a separate horizon, and this significantly limits the quality and effectiveness of the estimation (hereinafter referred to as prediction). Based on this it is easy to understand the opinion expressed by Grimble in 2002, the classic MVC theory performs large bandwidth, high gain, and implausibly high fluctuation in the control signal. Otherwise, the authors also encountered frequently this phenomenon, and we tried to limit this unacceptably large control signal change in an empirical way. After this, however, it became necessary to eliminate the unmanageable quality degradation (and the empirical tuning) caused by the neglected limitation in the design of the control algorithm. Therefore, the amendment proposed by Bitmead, Graves, and Wertz (1990) and Clarke and Mohtadi (1989), is very useful, being known in the literature as the Generalized Minimal Variance Control (GMVC) algorithm. In this method, the original objective function is supplemented with control signal constraints as well as the specified path. Naturally, this method also retained the one-step horizon prediction. Thus, the solution to reduce the indicated large control signal change could only be a significant weighting of the control signal or the change of the control signal in the objective function. This, in turn, reduces the control signal both outside and within the limits, and thus results in quality degradation. An effective solution to this problem is the Generalized Predictive Control (GPC) algorithm, developed by Clarke and Gawthrop. This in itself provides self-tuning adaptive control just like MVC and GMVC algorithms. However, - through the efficient constraint management algorithm and arbitrary extension of its prediction horizon - it also provides an effective solution to the listed deficiencies: handles the non-minimal phase and open-loop unstable processes as well as the unknown or variable dead time phenomena. Before superficially concluding that we have the ideal control algorithm (since the on-line estimation of the regression ARX model creates an opportunity to calculate the control algorithm in each sampling) let us look at the practical experience. In the case of application to linear or near-linear processes, this statement is

true. However, if we want to control a strongly nonlinear process, a linear ARX estimate will not always be adequate, especially if we need a horizon at a significant distance from the operation point. This problem can be solved by neural regression model estimation, which has been used effectively in many practical processes over the past two decades.

In the following sections, we will present in detail the practical implementation of the very popular generalized predictive control.

2. Thermal model of the indirect heated cathode assembly and comparison between MVC and GPC algorithms

2.1. Dynamic state-space model of the cathode heating process

The task is to implement a minimal variance (minimum standard deviation) adaptive control algorithm that controls the heating power of the indirectly heated massive cathode of a high-power electron beam machining equipment to a constant value. First, let us see what this task means. It is a known fact that in high power electron beam processing equipment – $n \times 100$ mA electron beam – an indirectly heated thermo-emission cathode is usually used as the electron source (**Figure 1**). Typically, the electron source structure consists of two parts: a massive cathode, which provides the machining current and a spiral shape primary cathode with an indirect heating function.

If the massive cathode is heated with a constant power - provided by electrons emitted by the primary cathode, operating in a saturated range - it is possible to achieve with a good approximation, a temperature at the emission surface of the active current cathode that is constant. The I_{nk} current, considered as the control signal passing through the primary cathode, heats the primary cathode to a T_n temperature, which as a result, emits a J_s electron beam of current density. The free electrons thus excited are accelerated by the constant potential U_s formed between the primary and the massive cathode, and the massive cathode is heated to Ts temperature. To operate efficiently, i.e. to minimize heat loss, the construction of the massive cathode also includes a thermal trap. That is, the diameter of the cathode narrows in the direction of the perception, 0.6 ... 0.7 mm, as opposed to the emission surface with a diameter of 1.4 mm. This solution leads to a reduction in

heat conduction, which also results in a reduction in the corresponding heat losses.

Naturally, this so-called "heat trap" must be imagined in the fixing direction of the primary cathode, thus reducing the power lost by the massive cathode due to heat conduction. Furthermore, all these result in nonlinearity in the characteristic of the co-structure of the cathode. The task is to design a minimum standard deviation (variance) MVC control algorithm, then a generalized minimum variance (GMVC), and finally a generalized predictive control (GPC). These algorithms consider the heating power of the massive cathode as an output. First, we must write the mathematical model describing the operation of the joint structure of the cathode, followed by the implementation of the MVC, GMVC, and GPC algorithms.

To determine the distribution of "massive cathode" heating power, it is necessary to know the potential distribution between toroid primary cathode (torus diameter R = 2.74 mm and wire diameter r = 0.15 mm) and the massive cathode (cylindrical with the following dimensions: l = 6.6 mm and 0.7 mm). From the motion of the electrons, we can determine the charge density distribution on the surface of the massive cathode, which is directly proportional to the incoming power density.

The electric current that heats the massive cathode – the current between the primary and massive cathode – travels along the lines of force of the electric field strength. To model the stationary mode – in the case of a constant cathode current or impact voltage – the current density and the heating power distribution on the surface of the massive cathode can be determined. Naturally, the effect of the current on the generated electrical potential can be determined iteratively over several cycles in succession. Based on all these, we used a global heating model by omitting the



Figure 1. Cross section of the indirectly heated cathode assembly and radiation heat transfer between the two cathodes.

steps of the electric field strength distribution and the electric current density approach. This global temperature can be determined based on equilibrium formulas of electrical powers. The heating power resulting from the collision of electrons emitted by the primary cathode, and accelerated by the stationary electric field between the cathodes, heats the interior of the massive cathode. Here, heat loss occurs on the surface through radiation and heat conduction. The heat loss through radiation is limited by the placement of certain cylindrical reflecting surfaces on the outer portion of the cathode assembly. The use of these also forms a kind of stability reserve by eliminating the sudden change of the emission current. Experimental results demonstrated the existence of positive feedback due to radiation. The surface of the massive cathode uses radiation to heat the primary cathode. It operates in saturated mode and emits a higher electron current density, which is proportional to the heating power.

To determine the dynamic state model the power of heat and the lost heat by radiation is required. Calculating the heat transfer surfaces for the radiation process, the equations of the thermal model is written under the following simplification conditions:

- The temperature distribution on the primer and massive cathode surfaces is uniform;
- The primary cathode heated by electric power, and by radiation; the surface distribution of the temperature is uniform. Heat losses are generated through radiation, thermal dissipation and convection heat losses appear only as unknown perturbations.
- The massive cathode is heated by the collision of electrons emitted by the primary cathode and accelerated by 800V. The losses occur through radiation and heat dissipation towards the thermal trap (thinned cross-section). Given the vacuum in the electron gun, no convection losses are expected, more precisely this member is considered an unknown perturbation member.
- Heat exchange through radiation is only relevant for the primary cathode and massive cathode, the primary cathode and thermal protector together with the massive cathode and thermal protector assemblies;
- The cylindrical concentric thermal protector partially reflects the effect of radiation, so that the protective surface also heats up. If the wall of this thermal protector is thin then the tem-

perature difference between the outer and inner walls is negligible.

The interaction created by radiation is the most important element in the heat transfer process. It is important to determine the interaction surfaces between the primary and massive cathodes and to determine the surfaces that completely lose heat during radiation. The primary cathode heats up under the action of a current considered as a controlled control signal I_{pk} . A portion of the radiant heat output emitted by the massive cathode corresponding to the massive angle reaching the massive cathode also heats the primary cathode (as positive feedback). This interaction can be followed in **Figure 1**.

Based on this, the primary cathode loses heat through radiation, which affects the energy required to heat it. Thermal heat conduction must be calculated on a massive cathode. Under these conditions, the following relation can give the lost heat output due to heat conduction:

$$P\lambda(T) = \lambda \cdot \pi \cdot \frac{d^2}{4} \left[8 \frac{\sigma r_0}{(\alpha r_0 + 1) \cdot \lambda \cdot d} T^{\alpha r_0 + 1} \right]^{\frac{1}{2}}$$
(1)

The dynamic mathematical model of the heating process can be given by the following differential equations:

$$\frac{d}{dt} \underbrace{\mathbf{x}}_{\mathbf{x}} = f(\underline{\mathbf{x}}, u, t) = \frac{d}{dt} \begin{bmatrix} Tp \\ Ts \end{bmatrix} = \\ \begin{bmatrix} \frac{1}{Cp} \begin{bmatrix} \Pr(Ts) \cdot S_{01} \cdot \varepsilon(Tp) + 2R \frac{\rho 0 \cdot Tp^{\alpha}}{r^2} I_{pk}^2 - \Pr(Tp) \cdot S_{00} \end{bmatrix} \\ \frac{1}{Cs} \begin{bmatrix} \Pr(Tp) \cdot S_{10} \cdot \varepsilon(Ts) + Pe(Tp, Us) - \Pr(Ts) \cdot S_{11} - P\lambda(Ts) \end{bmatrix} \end{bmatrix}$$
(2)

where:

- $-C_p$ and C_s are the heat capacity of the primer or massive cathode,
- $-P_r$ the heat quantity emitted by radiation and reflected by the cathodes, which depends on the temperature:
 - $P_r(T) = \alpha \cdot r_0 \cdot T^{\alpha r_0}$, where the reflection factor determining reflected heat power: $\varepsilon(T) = P_r(T)/(\sigma \cdot T^4)$,
- $-j_s(T_p, R(R-r))$ the saturated current density emitted by the primary cathode (which is heated by the current, which is the control signal),
- $-P_e(T_p, U_s)$) is the total electrical power that heats the massive cathode, this represents the measured output:

 $P_e(T_p, U_s) = 4 \cdot \pi \cdot R \cdot 10^4 \cdot v_j \cdot j_s(T_p, E(R-r)) \cdot U_s$

- $-P\lambda(T)$ is the loss of massive cathode power based on thermal heat dissipation.
- $-2 \cdot R \cdot r_0 \cdot T_p^{\alpha} \cdot r^2 I^2$ represents the heating power generated by the I_{pk} control current at the primary cathode.

Naturally, the heating takes place through the Joule effect, where certain corrections must also be taken into account due to the change in the temperature-dependent resistance of the cathode.

2.2. Generalized minimal variance control (GMVC) and generalized predictive control (GPC) algorithms

If the time constants and the implemented interfaces allow computer control for constant required power, it is important to limit the currents I_{pk} and I_s . In this case, the control block consists of the amplifiers required to measure I_{pk} , U_s and I_s and the phase control unit of the thyristors supplying the primary cathode. The measurement of the P_{MK} power and the limitation of I_{pk} and I_s currents can be solved within the numerical algorithm.

A comparative study was performed regarding the dynamic characteristics of a discrete tracking controller (prescribed T_s temperature) and a minimum standard deviation adaptive controller.

There are several estimation methods for determining linear regression model parameters. These methods can filter the measured noisy output signals, to estimate the number of degrees of the system models, and estimate the model parameters. These algorithms can be used in offline (non-recursive) and on-line (recursive, cyclic) mode. The resulting estimated models can be used to self-tune the controller parameters minimizing a given criterion function.

We assume that after estimation, we have obtained the following ARMA model type

$$P_{MK}^{\langle k \rangle} = \frac{B\left(z^{-1}\right)}{A\left(z^{-1}\right)} I_{pk}^{\langle k-\mu \rangle} + \alpha \frac{C\left(z^{-1}\right)}{A\left(z^{-1}\right)} e_{k}$$
(3)

where

- $-P_{MK}$ is the output signal,
- $-I_{nk}$ is the input control signal,
- $-\mu$ is the appropriate delay (expressed as a multiple of the dead time sampling time),
- $-B(z^{-1})/A(z^{-1})$ -) fraction is the transfer function of the plant,
- $-C(z^{-1})$ is polynomial characterizes the nature of noise. If this is a white noise then $C(z^{-1}) = 1$. The noise vector e_k is usually a normally dis-

tributed, zero-mean signal. The minimum variance controller performs at minimum of the following criterion:

$$J(u) = E\left\{ \left(P_{MK}^{\langle k+\mu+1 \rangle} - P_{MKpr} \right)^2 + \lambda \cdot \left(I_{pk}^{\langle k \rangle} - I_{pkpr} \right)^2 \right\}$$
(4)

where

- $-P_{MK}$ PMK is the controlled output,
- $-P_{MKpr}$ is the desired output (this value is modified as a function of the value of the difference between the maximum rated current of the electron beam and the estimated maximum current),
- $-I_{nk}$ is the control signal,
- $-I_{pkpr}^{PR}$ is the desired input value.

If the delay of the model is $\mu = 1$ 1and the parameters are known (polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ are given):

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2}$$

$$C(z^{-1}) = 1$$
(5)

The delay time and the degree of the polynomials were chosen experimentally (i.e., the smallest degree that still provides adequate accuracy). To determine the controller parameters, the following F(.) és G(.) polynomials need to be defined:

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-\mu}G(z^{-1})$$

$$F(z^{-1}) = 1, G(z^{-1}) = g_0 + g_1 \cdot z^{-1} + g_2 \cdot z^{-2}$$

$$1 = (1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2} + a_3 \cdot z^{-3}) \cdot 1 + z^{-1}(g_0 + g_1 \cdot z^{-1} + g_2 \cdot z^{-2})$$

The control signal is:

$$\begin{split} I_{pk}^{(k)} &= \frac{B(z^{-1})^2 F(z^{-1}) C(z^{-1}) + \lambda \cdot A(z^{-1}) C(z^{-1})}{B(z^{-1}) \cdot \left[B(z^{-1})^2 F(z^{-1})^2 + \lambda^2 C(z^{-1}) \right]} P_{MSpr} - \\ &- \frac{B(z^{-1}) F(z^{-1}) G(z^{-1})}{B(z^{-1})^2 F(z^{-1})^2 + \lambda^2 C(z^{-1})} P_{MS}^{(k)} \\ I_{pk}^{(k)} &= \frac{B(z^{-1})^2 + \lambda \cdot A(z^{-1})}{B(z^{-1}) \cdot \left[B(z^{-1})^2 + \lambda^2 \right]} P_{MSpr} - \\ &- \frac{B(z^{-1}) G(z^{-1})}{B(z^{-1})^2 + \lambda^2} P_{MS}^{(k)} \\ B(z^{-1}) \cdot \left[B(z^{-1})^2 + \lambda^2 \right] I_{pk}^{(k)} = \\ &= \left[B(z^{-1})^2 + \lambda \cdot A(z^{-1}) \right] P_{MSpr} - B(z^{-1})^2 G(z^{-1}) P_{MS}^{(k)} \end{split}$$

$$(7)$$

(6)

If the control signal is unconstrained $\lambda \rightarrow 0$, then the above equation will be simpler:

$$B\left(z^{-1}\right) \cdot I_{pk}^{\langle k \rangle} = P_{MKpr} - G\left(z^{-1}\right) P_{MK}^{\langle k \rangle}$$
(8)

Substituting this, we get the control law according to the MVC algorithm, such as:

In practice, the important fact is that the rounding errors of the DAC converter determining the control signal of the computer control system can be best modeled with an additive white noise, the correlation of which with the useful signal is negligibly small [4], [5]. The simulation includes two types of noise: a limited white noise at the output (that models the quantization errors) and a normally distributed noise, which affects the input measurements. Figure 2. shows the results of this GMVC control.

From this result we can observe that the classic GMVC, which does not include the limitation of either the control signal or the change of the control signal, generates a large control signal that swings the system. This is mainly seen where the values of heating power have high values and where the process characteristics are very steep. In this case, a 1mA change in the control signal results in a significant power change. Therefore, we can introduce an empirical control signal constraint performed by a filter. It can be seen that this stabilizes the process somewhat, although the ramp-up tracking error and the steady-state are not completely error-free either. It would be useful to formulate the constraint already in the objective function. The estimation error of the ARMAX model parameters of the estimated process can also be well observed (red line). Otherwise, this relationship is known, and can be used often in adaptive control algorithms (so-called for self-tuning control algorithms).

In the case of GPC control of Cathode Heating, the prediction is determined in advance for several horizons. It has already been mentioned that the GMVC algorithm can only be applied successfully if we limit the change of the control signal by some method. In the above case, an empirically introduced filter did this. Weakly limiting the absolute value of the control signal is not a good idea here because it penalizes the control current



Figure 2. GMVC control results without output filter (a) or with an output filter (b).

of 7-8 A at the steady-state and not its change. Therefore, we do not deal with this implementation. However, the GMVC algorithm for limiting the change of the control signal should be the same as the algorithm for GPC control (with one sampling horizon) without a strong constraint:

$$A_{\Delta} \left(z^{-1} \right) = \left(1 - z^{-1} \right) \left(1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} \right) =$$

= 1 + (a₁ - 1) z⁻¹ + (a₂ - a₁) z⁻² + (a₃ - a₂) z⁻³ - a₃ z⁻⁴
(10)

To determine the GPC algorithm the following equation is required:

$$\hat{y}_{k+p/k} = G_p(z^{-1}) \cdot y_k + F_p(z^{-1}) \cdot B_\mu(z^{-1}) \cdot \Delta u_{k+p-1}$$
(11)

We need to determine the polynomials $F(z^{-1})$ and $G(z^{-1})$ for different horizon values (*p*), as:

$$A_{\Delta}(z^{-1})F(z^{-1}) = 1 - z^{-p}G(z^{-1}) \Longrightarrow$$

$$(1 + (a_1 - 1)z^{-1} + (a_2 - a_1)z^{-2} + (a_3 - a_2)z^{-3} - a_3z^{-4}) \cdot F(z^{-1}) + z^{-p}G(z^{-1}) = 1$$

Case 1: p=1, n=3

$$\begin{split} F_1(z^{-1}) &= 1 + f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{\mu-1} z^{-p+1} = 1 \\ G_1(z^{-1}) &= g_0 + g_1 z^{-1} + \ldots + g_{n-1} z^{-n} = g_0^1 + g_1^1 z^{-1} + g_2^1 z^{-2} + g_3^1 z^{-3} \\ g_0^1 &= 1 - a_1, g_1^1 = a_1 - a_2, g_2^1 = a_2 - a_3, g_3^1 = a_3 \end{split}$$

Case 2: p=2, n=3 $F_2(z^{-1}) = 1 + f_1 z^{-1} + f_2 z^{-2} + \dots + f_{\mu-1} z^{-p+1} = 1 + f_1^2 z^{-1}$ $G_2(z^{-1}) = g_0 + g_1 z^{-1} + \dots + g_{n-1} z^{-n} = g_0^2 + g_1^2 z^{-1} + g_2^2 z^{-2} + g_3^2 z^{-3}$ $f_1^2 = 1 - a,$ $g_0^1 = f_1^2 \cdot (1 - a_1) + a_1 - a_2,$ $g_1^2 = f_1^2 \cdot (a_1 - a_2) + a_2 - a_3,$ $g_2^2 = f_1 \cdot (a_2 - a_3) + a_5, g_3^2 = f_1 \cdot a_3$

Case 3: p=3, n=3

$$F_3(z^{-1}) = 1 + f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{\mu-1} z^{-\mu+1} = 1 + f_1^3 z^{-1} + f_2^3 z^{-2}$$

 $G_3(z^{-1}) = g_0 + g_1 z^{-1} + \ldots + g_{n-1} z^{-n} = g_0^3 + g_1^3 z^{-1} + g_2^3 z^{-2} + g_3^3 z^{-3}$
 $f_1^3 = 1 - a_1,$
 $f_2^3 = f_1^3 \cdot (1 - a_1) + a_1 - a_2,$
 $g_0^3 = f_1^3 \cdot (a_1 - a_2) + f_2^3 \cdot (1 - a_1) + a_2 - a_3,$
 $g_1^3 = f_1^3 \cdot (a_2 - a_3) + f_2^3 \cdot (a_1 - a_2) + a_3,$
 $g_2^3 = f_1^3 \cdot a_3 + f_2^3 \cdot (a_2 - a_3),$

 $g_{2}^{3} = f_{2}^{3} \cdot a_{2}$

$$\begin{split} F_3(z^{-1}) &= 1 + f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{\mu-1} z^{-\mu+1} = 1 + f_1^4 z^{-1} + f_2^4 z^{-2} + f_3^4 z^{-3} \\ G_3(z^{-1}) &= g_0 + g_1 z^{-1} + \ldots + g_{n-1} z^{-n} = g_0^4 + g_1^4 z^{-1} + g_2^4 z^{-2} + g_3^4 z^{-3} \\ f_1^4 &= 1 - a_1, \\ f_2^4 &= f_1^4 \cdot (1 - a_1) + a_1 - a_2, \\ f_3^4 &= f_1^4 (a_1 - a_2) + f_2^4 (1 - a_1) + a_2 - a_3 \\ g_0^4 &= f_3^4 \cdot (1 - a_1) + f_2^4 \cdot (a_1 - a_2) + f_1^4 (a_2 - a_3) + a_3, \\ g_1^4 &= f_3^4 \cdot (a_1 - a_2) + f_2^4 \cdot (a_2 - a_3) + f_1^4 a_3, \\ g_2^4 &= f_2^4 \cdot a_3 + f_3^4 \cdot (a_2 - a_3), \\ g_3^4 &= f_2^4 \cdot a_3 \end{split}$$

Naturally, we can continue this and gradually extend the value of the horizon and we can write the following equation (for the value p = 1, 2, 3, 4 of the different horizon parameters) as:

$$\begin{split} \hat{y}_{k+1/k} &= \left(1 - a_1 + \left(a_1 - a_2\right)z^{-1} + \left(a_2 - a_3\right)z^{-2} + a_3z^{-3}\right) \cdot y_k + \\ 1 \cdot \left(b_0z^{-1} + b_1z^{-2} + b_2z^{-3}\right) \cdot \Delta u_k \\ \hat{y}_{k+2/k} &= \left(\int_1^2 \left(1 - a_1\right) + a_1 - a_2 + \left(\int_1^2 \left(a_1 - a_2\right) + a_2 - a_3\right)z^{-1} + \\ \left(f_1^2 \left(a_2 - a_3\right) + a_3\right)z^{-2} + f_1^2a_3z^{-3} \\ &+ \left(1 + \left(1 - a_1\right)z^{-1}\right) \cdot \left(b_0z^{-1} + b_1z^{-2} + b_2z^{-3}\right) \cdot \Delta u_{k+1} \end{split} \right)$$

$$\begin{split} \hat{y}_{k+3/k} &= \begin{pmatrix} f_1^{\,3}\left(a_1-a_2\right)+f_2^{\,3}\left(1-a_1\right)+a_2-a_3+\\ \left(f_1^{\,3}\left(a_2-a_3\right)+f_2^{\,3}\left(a_1-a_2\right)+a_3\right)z^{-1}+\\ +\left(f_1^{\,3}a_3+f_2^{\,3}\left(a_2-a_3\right)\right)z^{-2}+f_2^{\,3}a_3z^{-3} \end{pmatrix} \cdot y_k + \\ &+ \left(1+\left(1-a_1\right)z^{-1}+\left(f_1^{\,3}\left(1-a_1\right)+a_1-a_2\right)z^{-2}\right) \cdot \\ \left(b_0z^{-1}+b_1z^{-2}+b_2z^{-3}\right) \cdot \Delta u_{k+2} \\ & \left(f_3^{\,4}\left(1-a_1\right)+f_2^{\,4}\left(a_1-a_2\right)+f_1^{\,4}\left(a_2-a_3\right)+a_3+\\ +\left(f_3^{\,4}\left(a_1-a_2\right)+f_2^{\,4}\left(a_2-a_3\right)+f_1^{\,4}a_3\right)z^{-1}+\\ \left(f_2^{\,4}a_3+f_3^{\,4}\left(a_2-a_3\right)\right)z^{-2}+f_3^{\,4}a_3z^{-3} \end{pmatrix} \cdot y_k + \\ & \left(1+\left(1-a_1\right)z^{-1}+\\ \left(f_1^{\,4}\left(1-a_1\right)+a_1-a_2\right)z^{-2}+\\ \left(f_1^{\,4}\left(a_1-a_2\right)+f_2^{\,4}\left(1-a_1\right)+a_2-a_3\right)z^{-3} \right) \right) \cdot \end{split}$$

Rewriting this:



It can be seen here that the first equation cannot be used due to dead time. So, eliminating the first line from these matrices, and in this case, using the G_p , $FB^{<0>}$ and FB notations, the relevant equations for the free-response and the optimal control are:

$$\hat{Y}_{k}^{0} = \begin{bmatrix} \hat{y}_{k+2/k} \\ \hat{y}_{k+3/k} \\ \hat{y}_{k+4/k} \end{bmatrix} = G_{\mathbf{p}} \cdot \begin{bmatrix} y_{k} \\ y_{k-1} \\ y_{k-2} \\ y_{k-3} \end{bmatrix} + FB^{\langle 0 \rangle} \cdot \begin{bmatrix} \Delta u_{k-1} \\ \Delta u_{k-2} \\ \Delta u_{k-3} \end{bmatrix}$$

The value of the control signal sequence (unrestricted case) can be obtained such that:

$$\begin{bmatrix} \Delta u_k^* \\ \Delta u_{k+1}^* \\ \Delta u_{k+2}^* \end{bmatrix} = \left(F_B^T \cdot F_B + \lambda I \right)^{-1} \cdot F_B^T \left(R_k - \hat{Y}_k^0 \right)$$
(12)

where:

$$F_{B} = \begin{bmatrix} b_{0} & 0 & 0\\ (1-a_{1})b_{0}+b_{1} & b_{0} & 0\\ (f_{1}(1-a_{1})+a_{1}-a_{2})b_{0}+(1-a_{1})b_{1}+b_{2} & (1-a_{1})b_{0}+b_{1} & b_{0} \end{bmatrix}$$

Figure 3 shows the control results for the GPC algorithm without strong restriction (for different horizon values p=1, 2, 3). Here, the weight value of the control signal was chosen so that the heating current – as a control signal – does not reach the selected limit. In this case λ = 10.

From these results, it can be seen that even the one-step horizon can ensure stable operation, but the two- and three-step horizons already result in completely satisfactory operation. However, it is not enough to be able to control a given constant power. There is another condition, namely, whether or not the heated cathode can provide the machining current. This practically depends on two factors. First, we have to ask: whether the temperature of the massive cathode is high enough for the saturation current to be higher than the required machining current. This means that the process operates within the range defined by the space charge. Of course, it is not a solution if the temperature of the cathode is too high. In this case, the required current can indeed provide, but due to overheating, the life of the cathode is significantly reduced This is already a problem in terms of economy. It does not matter if we have to change cathodes on a monthly or daily basis. Secondly, the position of the cathode can also be a problem. A cathode inserted too deep already shuts off the jet current at a very low Wehnelt voltage. Thus, for full opening, the Wehnelt voltage is too low to ensure the formation of a "cross-over" focal point. So in this case, we cannot perform the machining itself either, because the radius focused in the plane of the machining means the "cross-over" mapping. However, by applying the results of modern control technology, both phenomena can easily be handled.

2.3. Implementation of the extended Kalman estimation

First, we look at the estimation of the saturation current. It involves two steps. We estimate the temperature of the solid cathode emission surface, and then assuming that the emission is provided by the so-called Schottky thermal emission, we can calculate the maximum available machining electron beam current, influenced by the saturation emission current.

The practical estimation of cathodes temperatures can be performed with the classical method of the extended-Kalman filter:







Figure 3. Cathode heating - GPC adaptive control for p = 1, 2, 3 step horizon.

- 1. In each sample are measured the control signal I_{PKk} and the output signal P_{MKk} sequence.
- 2. We determine the discrete state-space model matrices valid in the given step Φ_k and Γ_k . These can be determined by Taylor series of the non-linear equation (2) defining the cathode heating.
- 3. We determine the so-called deterministic (uncorrected) estimation, using a numerical approximation method (for example Euler or Runge-Kutta numerical methods):

$$\underline{x}_{k+1/k} = \underline{x}_{k/k} + \delta t \cdot \underline{f}\left(\underline{x}_{k/k}, \underline{u}_{k}\right)$$
(13)

In this case, some calculation errors can appear, as estimation errors, but the calculation will be much faster.

4. Calculation of the extrapolated standard deviation matrix (prediction step):

$$P_{k+1/k} = \Phi_k \cdot P_{k/k} \cdot \Phi_k^T + Q_k \tag{14}$$

5. Determination of the feedback gain vector (that provides the minimum standard deviation) :

$$K_{k+1} = P_{k+1/k} H_k^T \left(H_k P_{k+1/k} H_k^T + R_k \right)^{-1}$$
(15)

6. Calculation of the estimated state vector adjusted by feedback (updating step):

$$x_{k+1/k+1} = x_{k+1/k} + K_k \cdot \left(y_{k+1} - H_{k+1} \cdot x_{k+1/k} \right)$$
(16)

7. Determination of the feedback-corrected standard deviation matrix value (updating step):

$$P_{k+1/k+1} = \left[I - K_{k} \cdot H_{k}\right] P_{k+1/k}$$
(17)

8. Increase the number of cycle $k \rightarrow k + 1$ and jump to step 1.

In the case of cathode heating of the electron beam equipment, this estimation process can be followed in **Figure 4**.

We can see the estimation of the two temperatures, the estimated standard deviation of the massive cathode, and the estimated saturation current I_s (which is provided by Schottky cathode theory with a good approximation) [6], [7]. The confidence factor of the estimate is the change in the trace of the covariance matrices. A decrease in the relatively large initial value (P₁₁ = P₂₂ = 100) indicates a correct estimate. Two uncertainties are present in this task. First the uncertainty of the used model. This simplified model instead of the cathode temperature distribution, where a



Figure 4. Results of the cathode heating GPC control and extended Kalman filter estimation (temperature and saturation current).

uniform temperature was assumed for both cathodes. The second uncertainty, is the estimating algorithm. We did not use the very time-consuming numerical integration algorithms (Runge-Kutta, Adams, etc.) but the much simpler and faster Euler method where the Kalman filter compensates for the numerical inaccuracies.

The situation must occur at the time of t = 25 sec. Opening the valve between the electron gun and the machining chamber results in the phenomenon of convective heat transfer around the heated cathode due to the decreasing vacuum, which cools the cathode. The value of the available maximum saturation current begins to decrease significantly, which is somewhat compensated by the power regulator implemented by GPC, but not completely.

Furthermore, the resulting heating and then cooling cycles also affect the life of the cathode.

Therefore, it would be correct to design the required trajectory (heating rise path) in such a way that it also prolongs the life of the cathode itself. This can be achieved with optimal trajectory planning. In the next paragraph, we present its theoretical approach.

2.4. Optimal trajectory planning

In the case where we want to design the optimal trajectory to track - which is a completely different task than following the prescribed trajectory in real-time – linearization methods become meaningless. This task decides the required trajectory based on the entire trajectory, which should be the best based on the given objective function. Here we cannot fix a point wherein its immediate vicinity – would be worthwhile to apply the method of linearization. This is a completely different task than the control itself, but it is at least as important because it gives the stated goal of the best-prescribed trajectory.

If we know the mathematical model of the process, then the necessary conditions for optimization such as Hamilton-Jacobi or Hamilton-Jacobi-Bellman equations, can be easily formulated and described mathematically. Now there are several numerical methods for solving these. In many fields of applied sciences, we find different nonlinear dynamic optimal control problems with increasing complexity. Thus, following the theoretical approaches, several numerical algorithms have been developed and applied in the last few decades. Moreover, it can now be considered a separate area in science. **Figure 5.** shows the different numerical methods of optimal trajectory planning.

Based on these, we can see that currently optimal control, as a topic, means the application of the software that converts continuous nonlinear optimal control problems into a numerically solvable problem. Beginning in the 1980s, their main trend has been to transform into a nonlinear programming task (NLP) that has been solved by relatively well-studied methods. For example, the following software environments SNOPT [11], IPOPT [12] and KNITRO [13]. Originally, these methods were used to define the time scale of the solutions, so that they (the time values) were the same for all equations (state equations, the control signal, and auxiliary state equations). Convergence can be solved by compressing these points [14], [15]. Recent research has shown that direct orthogonal Gaussian quadrature yields good results. The point of this is that we choose not only

the weights but also the abscissa where we want to approach a function. The advantage of Gaussian quadrants is that for a given integer denoting the number of approximation terms, we always find an order w_j weight and an orthogonal $f_j(.)$ function so that the integral approximation approaching the optimal function of the optimal trajectory is as accurate as possible.

$$\int_{-1}^{1} f(x) \cdot dx = \sum_{i=1}^{n} w_i \cdot f_i(x)$$

This method is used by the methods known as "pseudo-spectrum" known as Legendre-Gauss, Legendre-Gauss-Radau, and Legendre-Gauss-Lobatto [16]. These methods make sense if the optimal trajectory to be determined and the differential equations describing the system make it necessary.

For general control problem formulation, a known general state model written for continuous system is usually used.

$$\frac{d}{dt}\underline{x}(t) = f(\underline{x}(t), \underline{u}(t), t)$$

$$\underline{y}(t) = h(\underline{x}(t), \underline{u}(t), t)$$
(18)

where $\underline{x}(t) \in \mathcal{H}^n$ is the vector, $\underline{y}(t) \in \mathcal{H}^p$ the output, and $\underline{u}(t) \in \mathcal{H}^m$ the control input vectors. It is also important to emphasize that this optimal control and the corresponding optimal trajectory calculation can also be used to generate the prescribed trajectory of the given system, along with which the controlled system meets certain prescribed conditions. The calculation of the generally valid control algorithm is only possible by directly considering the nonlinear characteristic. In the following, a method, which determines – in a more or less iterative way – the control signal valid in the whole range is presented.

The optimal control of the studied dynamic system means that it provides the minimum of the next objective function



Figure 5. Numerical methods of optimal trajectory planning.

$$J(\underline{u}) = \lambda \left(\underline{x}(tf)\right) + \int_{to}^{tf} L(\underline{x}(t), \underline{u}(t), t) dt$$
(19)

<u>The states x(t) and control signals u(t) of the system are not limited yet, the time horizon t_f of the control task is fixed and the target state $x(t_f)$ of the system is free. Following the known method, we determine the $H(\underline{x}, \underline{u}, \underline{p}, t)$ Hamiltonian function assigned to the problem as:</u>

$$H\left(\underline{x}(t),\underline{u}(t),\underline{p}(t),t\right) = \\ = L\left(\underline{x}(t),\underline{u}(t),p(t),t\right) + p^{T}\left(t\right) \cdot f\left(\underline{x}(t),\underline{u}(t),t\right)$$
(20)

The algorithm of the method is as follows:

- 1. In the interval $t \in [t_o, t_f]$ we choose an initial approximation of the control signal $u_o(t)$.
- 2. Using this, we compute a trajectory of the state equations on the time interval $t \in [t_0, t_p]$.
- 3. Using the boundary condition of the auxiliary state vector

$$\underline{p}(tf) = \frac{\partial \lambda(\underline{x}(tf))}{\partial \underline{x}},$$
(21)

we determine an approximation of the auxiliary state vector *p(t)* by integrating back in time the HJ equations.

4. We examine the following condition:

$$\left\|\frac{\partial H}{\partial \underline{u}}\right\| = \int_{0}^{t^{2}} \left[\frac{\partial H}{\partial \underline{u}}\right]^{T} \left[\frac{\partial H}{\partial \underline{u}}\right] dt \leq \varepsilon$$
(22)

5. If this is met, then we have obtained the optimal trajectory if it is not met, then we have to make the following correction:

$$\underline{u}(t)^{< k+1>} = \underline{u}(t)^{< k>} - S^* \cdot \frac{\partial H(x, p, u, t)}{\partial \underline{u}}$$
(23)

The iteration continues from the second step.

The question is what objective function we can choose. Practical observations prove that the lifetime of an indirectly heated cathode is determined by the lifetime of the primary cathode. If no unexpected event occurs, this service life of the cathode can be calculated based on the loss of mass or diameter due to its evaporation. If this mass loss value is denoted $M_w < g/cm^2 \cdot s >$ then according to [6] and [9] the estimated lifetime Δt can be calculated as the time interval during which the cathode loses Δm mass. We can write:

$$\Delta t = \frac{\Delta m}{M_w(T) \cdot S_{PK}} = \frac{\gamma_w \cdot \Delta V}{M_w(T) \cdot \pi \cdot \eta_i \cdot (2r) \cdot l_{PK}}$$

where $\gamma_w = 19.5 \text{ g/cm}^3$ is the density of the tungsten cathode, l_{p_K} <cm> the helical length of the primary cathode, and assuming that the diameter of the primary cathode decreases from $\eta_i(2r)$ to $\eta_{i+1}(2r)$ then:

$$\Delta V = \frac{\pi}{4} l_{PK} \cdot \left(2r\right)^2 \cdot \left(\eta_i^2 - \eta_{i+1}^2\right)$$

result:

$$\Delta t = \frac{\gamma_{w} \cdot (2r) \cdot (\eta_{i}^{2} - \eta_{i+1}^{2})}{4M_{w}(T) \cdot \eta_{i}}$$

The weight loss M_{w} is strongly temperature-dependent (see Figure 6),) therefore one solution would be to calculate the percentage of the cathode that evaporates at a constant controlled temperature for each switch-on and machining, then the sum of these should be the smallest projected over a given time interval. This results in the cathode temperature being kept as low as possible, but not below a certain lower value. We obtain almost the same result assuming the following empirical result [8]. The performance of a tungsten direct-heated thermal emission cathode is proportional to the U power of the supply voltage U^{1.6}, the luminance is proportional to the power output U^{3.4,} and its lifetime is proportional to the power U⁻¹⁶ [10].

3. Conclusions

The dissertation studies processing equipment that provides industry-proven technology under today's knowledge, which is a serious alternative to laser machining. The machining itself is more accurate, homogeneous, and more resistant than



Figure 6. Cathode weight loss as function of temperature.

laser machining because the machining itself takes place under vacuum. This is electron beam machining, which has proven its suitability in the last half-century in the fields of welding, surface heat treatment, evaporation, melting, and currently, 3D printing with additive has made a significant breakthrough. However, all this can only be achieved if the parameters of the process itself are properly controlled. Such is the cathode assembly that provides the jet current itself. It is important to control the appropriate safe and accurate emission surface temperature, as well as to know the saturation (maximum) jet stream, and to provide a mode of operation that provides maximum life for this assembly. For a given existing configuration, we have demonstrated that an adaptive GPC control can provide stable operation, while an extended Kálmán filter can adequately provide a preliminary estimate of cathode surface temperature and saturation beam current, which makes the technology safe. In the dissertation, we supplemented all this with an optimal trajectory calculation that ensures maximum service life, which exerts its effect with the help of the prescribed power. All of this was developed to improve the performance of a CTW 5/60 electron gun, which is being refurbished at Sapientia Hungarian University of Transylvania.

References

- Åström K. J., Wittenmark B.: On self-tuning regulators. Automatica, 9. (1973) 185–199.
- [2] Keviczky L., Hetthéssy J., Hilger M., Kolostori J.: Self-tuning adaptive control of cement raw material blending. Automatica, 1978.
- [3] Methods of Model Based Process Control in Proceedings of the NATO Advanced Study Institute on Methods of Model Based Process Control. Antalya, Turkey, August 7–17, 1994.
- [4] Bennett W. R.: *Spectra of Quantized Signals*. Bell System Technical Journal, 27. (1948) 446–472.
- [5] Gray R. M., Neuhoff D. L.: *Quantization*. IEEE Transactions on Information Theory, IT-44/6. (1998) 2325–2383.

- [6] Hinrics C. H., Mackie W. A., Pincosy P. A., Poulsen P.: The Extended Schottky Cathode. IEEE Transactions on Electron Devices, 37/2. (1990) 2575–2580.
- [7] Hartman W., Kirkman G., Dominic V., Gundersen M. A.: Super Emissive Self-Heated Cathode for High Power Applications. IEEE Transactions on Electron Devices, 36/4. (1989) 825–826.
- [8] Plante E. R., Sessoms A. B.: Vapor Pressure and Heat of Sublimation of Tungsten. Journal of Research of the National Bureau of Standards, Physics and Chemistry, 77A/2. March-April, 1973.
- [9] Virag M., Murín J.: Thermal Field Simulation of a Tungsten Filament Lamp Referring to its Lifetime. Journal of Electrical Engineering, 56/ 9-10. (2005) 252–257.
- [10] Özbek H., Çil C., Rodoplu A.: Effect of Architecture and Operating Conditions of Vehicle Bulb Lifetime in Automotive. International Journal of Mechanical and Mechatronics Engineering, 13/6. (2019)
- [11] Gill P. E., Murray W., Saunders M. A.: SNOPT: An SQP Algorithm for Large Scale Optimization. SIAM Review, 47/1. (2005) 99–131.
- [12] Wächter A., Biegler L.T.: On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming. Mathematical Programming, 106. (2006) 25–57. https://doi.org/10.1007/s10107-004-0559-y
- [13] Byrd R. H., Nocedal J., Waltz R. A.: Knitro: An Integrated Package for Nonlinear Optimization. In: Di Pillo G., Roma M. (Eds.): Large Scale Optimization, Nonconvex Optimization and Its Applications, 83. (2006) 35–59.
- [14] Betts J. T.: Practical methods for optimal control and estimation using nonlinear programming. 2nd edition, 2010, SIAM.
- [15] Zhao Y., Tsiotras P.: Density Functions for Mesh Refinement in Numerical Optimal Control. Journal of Guidance, Control and Dynamics, 34/1. (2011). https://doi.org/10.2514/1.45852
- [16] Garg D., Patterson M. A., Hager W. W., Rao A. V., Benson D., Huntington G. T.: An overview of three pseudospectral methods for numerical Solution of Optimal Control. Advances in the Astronautical Sciences, 135. 2009.