META-CONTROL OF AN OPTIMIZATION ALGORITHM

WOLF KOHN, ZELDA B. ZABINSKY, VLADIMIR BRAYMAN
HYNOMICS CORPORATION
UNIVERSITY OF WASHINGTON, INDUSTRIAL ENGINEERING

Abstract. Optimization algorithms usually rely on the setting of parameters, such as barrier coefficients. We have developed a generic meta-control procedure to optimize the behavior of given iterative optimization algorithms. In this procedure, an optimal continuous control problem is defined to compute the parameters of an iterative algorithm as control variables to achieve a desired behavior of the algorithm (e.g., convergence time, memory resources, and quality of solution). The procedure is illustrated with an interior point algorithm to control barrier coefficients for constrained nonlinear optimization.

1. Optimization Problem. Optimization algorithms often include parameters that affect the overall performance of the algorithm. For example, interior point methods rely on barrier parameters. While it is shown that the algorithm will converge to optimality as the barrier parameters converge to zero, the specific values of the parameters are chosen empirically [6, 7]. We have developed an approach that can choose barrier values that not only guarantees convergence, but does so as fast as possible. Moreover, our approach is very general. In this paper, we present a control theoretic approach that allows us to analytically prescribe the parameter values of a generic algorithm that optimizes the performance of the algorithm. A very general measure of performance is used in our methodology, and some examples include the rate of convergence, memory usage, and quality of solution. We use the example of barrier parameters in an interior point algorithm with a Newton type descent direction to illustrate how to construct an optimal control problem that optimizes the performance of the algorithm by controlling the values of the parameters.

Our approach is to create an optimal control problem which is based on the original problem to be optimized as well as the algorithmic criterion. This optimal control formulation leads to a system of first order differential equations that constitute the necessary conditions for optimality. The solution to this system of differential equations provides not only the optimal solution to the original problem, but also the parameter values that optimize the performance of the algorithm. While it is difficult to directly solve this system of differential equations, because it is a two-point boundary value problem, we construct an associated variational problem with a solution that coincides with the optimal control problem. We construct the Lagrangian for this variational problem from the system of first order differential equations using the inverse Lagrangian technique. Finally, the variational problem is easily solved using direct methods.

We begin by formulating a standard (static) optimization problem, with $\mathcal{N}$ decision variables, $\mathcal{M}_1$ inequality constraints and $\mathcal{M}_2$ equality constraints. The original problem (P1) is

$$\begin{aligned} &\text{minimize} && \tilde{C}(x) && && \text{(P1)} \\ &\text{subject to} && \tilde{g}_i(x) \geq 0 && \text{for } i = 0, ..., \mathcal{M}_1 - 1 \\ & && \widetilde{h}_j(x) = 0 && \text{for } j = 0, ..., \mathcal{M}_2 - 1 \end{aligned}$$

where $x = (x_0, x_1, ..., x_{\mathcal{N}-1})^T \in \mathcal{R}^{\mathcal{N}}$ is the vector of decision variables. We assume the objective function $\tilde{C}(x)$ and all of the constraints $\tilde{g}_i(x)$ and $\widetilde{h}_j(x)$ are at least once continuously differentiable.

An iterative optimization algorithm for solving problem (P1) can be generically expressed as the following iteration,

$$\tilde{x}_{k+1} = \tilde{x}_k + \tilde{f}\left(\tilde{x}_k, \tilde{u}_k\right) \tag{1.1}$$

for $k = 0, 1, \ldots,$ where the function $\tilde{f}$ specifies the algorithm and maps $\mathcal{R}^{\mathcal{N}} \times U$ to $\mathcal{R}^{\mathcal{N}}$, with parameters $\tilde{u}_k \in U$, the set of feasible parameter values for the optimization algorithm [3]. We assume an initial point $\tilde{x}_0$ is given We also assume that the iterative equation (1.1) specified by $\tilde{f}$ converges, i.e., $\lim_{k\to\infty} \tilde{x}_k = \bar{x}$.

Given the specifics of the optimization algorithm with $\tilde{f}\left(\tilde{x}_k, \tilde{u}_k\right)$, we wish to control the parameters $\tilde{u}_k$ to achieve a desired behavior. One example of such behavior is to have the optimization algorithm converge as fast as possible. An example of the function $\tilde{f}\left(\tilde{x}_k, \tilde{u}_k\right)$ used later in the paper is a Newton type descent search with barrier parameters $\tilde{u}_k$. Typically the barrier parameters $\tilde{u}_k$ are chosen empirically, but we give a formal procedure for determining $\tilde{u}_k$ to achieve the desired behavior of the algorithm.

The strategy we use is to develop a *meta-control* procedure based on a continualization of the iteration [13]. In the continualization procedure, we convert iterations of the form (1.1) to a controlled differential equation in terms of $x(\tau)$ and $u(\tau)$. The variable $\tau \geq 0$, is a continuous convergence parameter corresponding to the discrete convergence parameter k. The differential equation is of the form,

$$\frac{dx(\tau)}{d\tau} = f(x(\tau), u(\tau)) \tag{1.2}$$

for all $u(\tau) \in U$, where $U \subseteq \mathcal{R}^{\mathcal{M}}$ is the set of feasible parameter values for the optimization algorithm. Later, in section 2, we introduce $\hat{u}(\tau, h(\tau)) = u(\tau)$ to emphasize that our computational approach includes two related convergence parameters (τ and $h(\tau)$) for u.

We define f in (1.2) to reflect the iteration characterized by $\tilde{f}$, by rewriting the iteration (1.1) as

$$\tilde{x}_{k+1} = \tilde{x}_k + \Delta_k f\left(\tilde{x}_k, \tilde{u}_k\right) \tag{1.3}$$

with $\Delta_k > 0$, the $\lim_{k\to\infty} \Delta_k = 0$, $\sum_{k=0}^{\infty} \Delta_k = \infty$, and $f\left(\tilde{x}_k, \tilde{u}_k\right) = \tilde{f}\left(\tilde{x}_k, \tilde{u}_k\right)/\Delta_k$. We assume the function f is Lipschitz continuous with Lipschitz constant less than one (contraction mapping) [13]. Since the original algorithm specified by $\tilde{f}$ is assumed to converge, i.e., $\lim_{k\to\infty} \tilde{x}_k = \bar{x}$, the continualization process ensures that the limit point of $x(\tau)$ equals the optimal solution in the original domain, i.e., $\lim_{\tau\to\infty} x(\tau) = \lim_{k\to\infty} \tilde{x}_k$. We formalize this with the following theorem.

THEOREM 1.1. *Let $\{\tilde{x}_k\}$ be a sequence in $\mathcal{R}^{\mathcal{N}}$ generated by $\tilde{x}_{k+1} = \tilde{x}_k + \tilde{f}\left(\tilde{x}_k, \tilde{u}_k\right)$, for a sequence $\{\tilde{u}_k\}$, that converges, $\lim_{k\to\infty} \tilde{x}_k = \bar{x}$. Also, let $f\left(\tilde{x}_k, \tilde{u}_k\right)$ be a contraction mapping from a subset of $\mathcal{R}^{\mathcal{N}} \times U$ into $\mathcal{R}^{\mathcal{N}}$, as above. Then there exists $x(\tau)$ satisfying $\frac{dx(\tau)}{d\tau} = f(x(\tau), u(\tau))$ such that $\lim_{\tau\to\infty} x(\tau) = \lim_{k\to\infty} \tilde{x}_k$.*

The proof can be found in [14, Theorem 2.3.1] and [15, Theorem 2.1].

The relationship between a discrete iterative procedure, as in (1.1), and a continuous differential equation, as in (1.2), is also discussed briefly in McCormick [16, pg.s 143-147]. An analogy is made between minimizing a function and rolling a boulder down the side of a mountain. The trajectory of the boulder would approximately satisfy the differential equation in terms of the gradient of the function. This analogy can be used to motivate Cauchy's method of steepest descent, as well as the classical version of Newton's method. While most nonlinear programming methods use the discrete iterative approach, we concentrate on working with the differential equation (1.2) directly.

The approximation of an iterative process with a differential equation provides several key advantages, in addition to the determination of the algorithm parameters $u(\tau)$:

1. We can formulate the numerical solution of (1.2) using higher order integration methods [2],
2. We can use stability analysis tools developed for differential equations,
3. We can apply the techniques of differential geometry to solve the differential equation,
4. We may take advantage of sophisticated computer algebra tools for analysis and more efficient computation.

Numerical and symbolic methods for solving differential equations are well-developed (see for example [17]), and are used in our meta-control computational schema. We are now ready to define the optimal control problem.

2. Optimal Control of the Algorithm. The optimal control problem (P2) is formulated as follows,

$$\min_{\substack{u(\tau)\\ \tau\in[0,T)}} \quad \int_0^T \phi(x(\tau),u(\tau))d\tau + \Psi\left(x(T)\right) \qquad \text{(P2)}$$

$$\text{subject to} \quad \frac{dx(\tau)}{d\tau} = f(x(\tau),u(\tau)) \qquad \text{for } \tau\in[0,T)$$

$$u(\tau)\in U \subseteq \mathcal{R}^{\mathcal{M}} \qquad \text{for } \tau\in[0,T)$$

with state vector $x(\tau)\in\mathcal{R}^{\mathcal{N}}$ and $x(0)$ given. The control vector is $u(\tau)\in U\subseteq\mathcal{R}^{\mathcal{M}}$. The functions ϕ and Ψ are chosen by the algorithm designer, in order to achieve objectives such as minimizing convergence time, minimizing memory usage, or maximizing the quality of the solution (minimizing solution error). In section 4, an example is presented specifying ϕ and Ψ to minimize the time for convergence. The function f represents the algorithm, as given in (1.2).

The optimal control problem (P2) can be solved by constructing and solving the necessary conditions for optimality. These necessary conditions for optimality are given by the minimum principle of Pontryagin [1]. The necessary conditions utilize the Hamiltonian of the system (P2), which is defined as

$$\mathcal{H}\left(x(\tau),u(\tau),p(\tau)\right) = \phi(x(\tau),u(\tau)) + p(\tau)^T f(x(\tau),u(\tau)),$$

where $p(\tau)$ is the costate associated with the problem [18]. In order for the control $u^*(\tau)$ to be optimal with $x^*(\tau)$ a corresponding trajectory, there must exist $p(\tau)\in\mathcal{R}^{\mathcal{N}}$ such that the following necessary conditions are satisfied:

$$\frac{dx^*(\tau)}{d\tau} = \left(\frac{\partial\mathcal{H}\left(x^*(\tau),u^*(\tau),p^*(\tau)\right)}{\partial p}\right)$$

$$\frac{dp^*(\tau)}{d\tau} = -\left(\frac{\partial\mathcal{H}\left(x^*(\tau),u^*(\tau),p^*(\tau)\right)}{\partial x}\right)^T$$

$$p(T) = \frac{\partial \Psi (x(T))}{\partial x},$$

$$x^* (0) = x_0$$

$$u^* (\tau) \in U (\tau)$$

for $\tau \in [0, T)$, where the Hamiltonian function has an absolute minimum as a function of u at $u^* (\tau)$, that is

$$\mathcal{H} (x^*(\tau), u^*(\tau), p^*(\tau)) \leq \mathcal{H} (x^*(\tau), u(\tau), p^*(\tau)), \tag{2.1}$$

for all $u (\tau) \in U (\tau)$, $\tau \in [0, T)$.

In order to solve the necessary condition on $u (\tau)$ in (2.1), we define an iterative procedure to minimize the Hamiltonian. We construct a convergence sequence $\{\breve{u}_s(\tau)\}$ generated by

$$\breve{u}_{s+1}(\tau) = \breve{u}_s(\tau) + \breve{W} (x(\tau), \breve{u}_s(\tau), p(\tau)) \tag{2.2}$$

such that

$$\lim_{s \to \infty} \breve{u}_s(\tau) = u^* (\tau), \text{ when } x(\tau) = x^*(\tau), \ \tau \in [0, T).$$

The algorithm defining the iterative step in (2.2) is specified by $\breve{W} (x(\tau), \breve{u}_s(\tau), p(\tau))$. An example of $\breve{W} (x(\tau), \breve{u}_s(\tau), p(\tau))$ based on a Newton step is discussed in section 4, and equation (4.5).

The key observation is: for each $\tau \in [0, T)$, (2.2) looks like (1.1) with $\breve{u}_s(\tau)$ analogous to $\tilde{x}_k$ and $\breve{W}$ analogous to $\tilde{f}$. We again apply the continualization procedure used to develop (1.2) from (1.1). Now we continualize (2.2) by introducing σ as a continuous parametrization of s, analogous to τ parameterizing k. We interpret σ as a convergence parameter, and, in developing the meta-control approach, we assume that σ is functionally related to τ, i.e. $\sigma = h (\tau)$, where h is a given continuous function. For example, if $\sigma = \tau/10$, then this indicates that the computation of $\breve{u}_{s+1}$ is embedded in the computation of $x (\tau)$. In principle, we could introduce a new variable $\hat{u} (\tau, \sigma)$, which is a continualized version of $\breve{u}_s(\tau)$; however, σ is related to τ ($\sigma = h (\tau)$), so we can express the continualized variable as a function of τ only, $\hat{u} (\tau, h (\tau)) = u (\tau)$.

The continuous version of (2.2) yields

$$\frac{du (\tau)}{d\tau} = W (x(\tau), u (\tau), p(\tau)),$$

where $u (\tau) = \hat{u} (\tau, h (\tau))$ is a continualized version of $\breve{u}_s(\tau)$ and $W (x(\tau), u (\tau), p(\tau))$ is a continualized version of $\breve{W} (x(\tau), \breve{u}_s(\tau), p(\tau))$ [13]. As in (1.3), we define W as follows,

$$W (x(\tau), u (\tau), p(\tau)) = \frac{1}{\delta_s} \breve{W} (x(\tau), \breve{u}_s(\tau), p(\tau)) \frac{dh (\tau)}{d\tau}$$

where δ_s is analogous to Δ_k in (1.3) with similar assumptions. Since $\lim_{s \to \infty} \breve{u}_s(\tau) = u^* (\tau)$, we conclude, by theorem 1.1, that $\lim_{\sigma \to \infty} \hat{u} (\tau, \sigma) = u^* (\tau)$.

We next summarize our approach. To solve the original optimization problem (P1), we solve three coupled differential equations, $\tau \in [0, T)$,

$$\frac{dx(\tau)}{d\tau} = f(x(\tau), u(\tau)) \tag{2.3}$$

$$\frac{du(\tau)}{d\tau} = W(x(\tau), u(\tau), p(\tau)) \tag{2.4}$$

$$\frac{dp(\tau)}{d\tau} = -\left(\frac{\partial \mathcal{H}(x(\tau), u(\tau), p(\tau))}{\partial x}\right)^T \tag{2.5}$$

with boundary conditions

$$\begin{aligned} x(0) &= x_0, \quad u(0) = u_0, \\ p(T) &= \frac{\partial \Psi(x(T))}{\partial x} \end{aligned} \tag{2.6}$$

given. Under mild smoothness assumptions [11], the solution to this set of equations, $x(\tau)$ evaluated at $\tau = T$, i.e. $x(T)$, closely approximates the original optimal solution $\bar{x}$.

3. Inverse Lagrangian. In this section we discuss a solution approach to solving the system of differential equations in (2.3) - (2.5). The primary difficulty in solving this system is that we have a two-point boundary value problem with initial conditions for equations (2.3) and (2.4) given at $\tau = 0$, and the terminal condition for equation (2.5) given at $\tau = T$. It is well known that this is a difficult problem to solve [2]. Our solution approach is to convert this two-point boundary value problem into a variational problem such that the solution to the variational problem coincides with the solution to the original system (2.3) - (2.5). Typically, the solution to a variational problem is obtained by solving a differential equation called the Euler-Lagrange equation [10]. We use the inverse Lagrangian approach [5], where we start with a differential equation, and derive a variational problem.

More precisely, we want to formulate a variational problem (P3) of the form

$$\min_{\substack{z(\tau) \\ \tau \in [0,T)}} \int_0^T L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right) d\tau \tag{P3}$$

with a Lagrangian $L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ that is twice continuously differentiable in the first two of its arguments. The solution to this problem is given by solving suitably constructed Euler-Lagrange equations [10]

$$\frac{dL_v(z(\tau), v(\tau), \tau)}{d\tau} - L_z(z(\tau), v(\tau), \tau) = 0, \tag{3.1}$$

where L_z is the derivative of L with respect to its first argument, and L_v is the derivative of L with respect to its second argument. Notice that equation (3.1) is a second order equation in terms of $z(\tau)$ because, by definition, $v(\tau) = \frac{dz(\tau)}{d\tau}$. Expanding the derivative in equation (3.1) and rearranging terms yields,

$$\frac{d^2 z(\tau)}{d\tau^2} = L_{vv}^{-1}(z(\tau), v(\tau), \tau)\left(L_z(z(\tau), v(\tau), \tau) - L_{vz}(z(\tau), v(\tau), \tau)\frac{dz(\tau)}{d\tau}\right), \tag{3.2}$$

where we assume the inverse $L_{vv}^{-1}(z(\tau), v(\tau), \tau)$ exists.

We now construct the appropriate Euler-Lagrange equation from the original system (2.3) - (2.5) and find a corresponding Lagrangian. We start by defining z as a vector-valued function, z: $[0, T) \rightarrow \mathcal{R}^{2\mathcal{N}+\mathcal{M}}$ as follows

$$z(\tau) = \begin{bmatrix} x(\tau) \\ u(\tau) \\ p(\tau) \end{bmatrix}, \quad \tau \in [0, T). \tag{3.3}$$

In order to bring the original system of equations (2.3) - (2.5) into the form of equation (3.2), we take a derivative with respect to τ on both sides of (2.3) - (2.5) to get,

$$\frac{d^2 z(\tau)}{d\tau^2} = \Theta\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right), \tag{3.4}$$

where the vector function $\Theta\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ is

$$\Theta\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right) = \begin{bmatrix} f_x(x(\tau), u(\tau)) \frac{dx(\tau)}{d\tau} + f_u(x(\tau), u(\tau)) \frac{du(\tau)}{d\tau} \\ W_x(x(\tau), u(\tau), p(\tau)) \frac{dx(\tau)}{d\tau} + W_u(x(\tau), u(\tau), p(\tau)) \frac{du(\tau)}{d\tau} \\ \quad + W_p(x(\tau), u(\tau), p(\tau)) \frac{dp(\tau)}{d\tau} \\ -\frac{\partial^2 \mathcal{H}(x(\tau),u(\tau),p(\tau))}{\partial x^2} \frac{dx(\tau)}{d\tau} - \frac{\partial^2 \mathcal{H}(x(\tau),u(\tau),p(\tau))}{\partial x \partial u} \frac{du(\tau)}{d\tau} \\ \quad -\frac{\partial^2 \mathcal{H}(x(\tau),u(\tau),p(\tau))}{\partial x \partial p} \frac{dp(\tau)}{d\tau} \end{bmatrix} \tag{3.5}$$

with boundary conditions

$$z(0) = \begin{bmatrix} I_{\mathcal{N}\times\mathcal{N}} & 0 & 0 & 0 \\ 0 & I_{\mathcal{M}\times\mathcal{M}} & 0 & 0 \\ 0 & 0 & 0_{\mathcal{N}\times\mathcal{N}} & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ u_0 \\ p(0) \end{bmatrix} \tag{3.6}$$

$$z(T) = \begin{bmatrix} 0_{\mathcal{N}\times\mathcal{N}} & 0 & 0 & 0 \\ 0 & 0_{\mathcal{M}\times\mathcal{M}} & 0 & 0 \\ 0 & 0 & I_{\mathcal{N}\times\mathcal{N}} & 0 \end{bmatrix} \begin{bmatrix} x(T) \\ u(T) \\ \frac{\partial \Psi(x(T))}{\partial x} \end{bmatrix}. \tag{3.7}$$

We assume the following smoothness for f, W and $\mathcal{H}$. We assume, for all τ, that f is twice jointly differentiable with respect to x and u, and W is twice jointly differentiable with respect to x, u, and p. Moreover, we assume $\mathcal{H}$ is jointly continuously differentiable three times with respect to x, u, and p. These assumptions imply that Θ is twice jointly differentiable with respect to its first two arguments.

In the following proposition, we define $L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$, a Lagrangian that is twice continuously differentiable in its first two arguments. The Euler-Lagrange equations associated with $L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ are (3.4)-(3.7). If the Lagrangian defined in the proposition is used in the variational problem (P3), then the solution to the variational problem gives a solution to (3.4) with boundary conditions (3.6), (3.7).

PROPOSITION 3.1. *Consider the variational problem (P3) with the Lagrangian* $L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ *and associated Euler-Lagrange equation (3.4). Then the Hessian of the Lagrangian given by* $L_{v_i v_k}\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right) = \Phi_{ik}$, *for* $i, k = 1, \ldots, 2\mathcal{N} + \mathcal{M}$, *satisfies*

$$2\frac{d\Phi_{ik}}{d\tau} + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left(\Phi_{ij} (\Theta_j)_{v_k} + \Phi_{kj} (\Theta_j)_{v_i}\right) = 0 \tag{3.8}$$

along the trajectories generated by (3.4)-(3.7), with $v_i = \frac{dz_i(\tau)}{d\tau}$ *and* $(\Theta_j)_{v_k}$ *denoting the partial derivative of* Θ_j *with respect to* v_k.

Proof. We derive a smooth Lagrangian $L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ using $\Theta\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ as defined in equation (3.4). First we rewrite equation (3.4) by introducing a new variable $v(\tau)$ as follows,

$$\begin{aligned} \frac{dz_i(\tau)}{d\tau} &= v_i(\tau) \\ \frac{dv_i(\tau)}{d\tau} &= \Theta_i\left(z(\tau), v(\tau), \tau\right) \end{aligned} \tag{3.9}$$

for $i = 1, \ldots, 2\mathcal{N} + \mathcal{M}$. Our approach to find $L\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ consists of finding an expression for its Hessian, $\{L_{v_i v_k},\, i, k = 1, \ldots, 2\mathcal{N} + \mathcal{M}\}$ in terms of the given functions $\{\Theta_i\left(z(\tau), v(\tau), \tau\right),\, i = 1, \ldots, 2\mathcal{N} + \mathcal{M}\}$ and some of their derivatives. Then we construct a particular Lagrangian from this Hessian using a simple quadrature.

Expanding (3.1) we obtain,

$$\sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ L_{v_i z_j}(z(\tau), v(\tau), \tau) v_j(\tau) + L_{v_i v_j}(z(\tau), v(\tau), \tau) \frac{dv_j(\tau)}{d\tau} \right\}$$
$$+ L_{v_i \tau}(z(\tau), v(\tau), \tau) - L_{z_i}(z(\tau), v(\tau), \tau) = 0$$

for $i = 1, \ldots, 2\mathcal{N} + \mathcal{M}$ with $v_j(\tau)$ and $\frac{dv_j(\tau)}{d\tau}$ given by the corresponding expressions in (3.9).

For the purposes of simplifying expressions, we drop the arguments $\left(z(\tau), \frac{dz(\tau)}{d\tau}, \tau\right)$ in the expressions. Then substituting Θ_j for $\frac{dv_j(\tau)}{d\tau}$, we obtain,

$$\sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ L_{v_i z_j} v_j + L_{v_i v_j} \Theta_j \right\} + L_{v_i \tau} - L_{z_i} = 0$$

for $i = 1, \ldots, 2\mathcal{N} + \mathcal{M}$. Differentiating both sides of these equations with respect to v_k, $k = 1, \ldots, 2\mathcal{N} + \mathcal{M}$, and using $(\Theta_j)_{v_k}$ to denote the partial derivative of Θ_j with respect to v_k, we obtain,

$$\sum_{j=1}^{2\mathcal{N}+\mathcal{M}} L_{v_i z_j v_k} v_j + L_{v_i z_k} + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ L_{v_i v_j v_k} \Theta_j + L_{v_i v_j} (\Theta_j)_{v_k} \right\} + L_{v_i \tau v_k} - L_{z_i v_k} = 0 \tag{3.10}$$

for $i, k = 1, \ldots, 2\mathcal{N} + \mathcal{M}$. By exchanging i with k in (3.10), we can write the symmetric expression:

$$\sum_{j=1}^{2\mathcal{N}+\mathcal{M}} L_{v_k z_j v_i} v_j + L_{v_k z_i} + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ L_{v_k v_j v_i} \Theta_j + L_{v_k v_j} (\Theta_j)_{v_i} \right\} + L_{v_k \tau v_i} - L_{z_k v_i} = 0 \tag{3.11}$$

for $i, k = 1, \ldots, 2\mathcal{N} + \mathcal{M}$. Adding (3.10) and (3.11) term by term and using a smoothness condition on L, we can eliminate terms that do not involve L_{vv} in some manner, and obtain the following set of symmetric equations:

$$\sum_{j=1}^{2\mathcal{N}+\mathcal{M}} 2L_{v_i v_k z_j} v_j + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} 2L_{v_i v_k v_j} \Theta_j + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} L_{v_i v_j} \left(\Theta_j\right)_{v_k} +$$
$$+ \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} L_{v_k v_j} \left(\Theta_j\right)_{v_i} + 2L_{v_i v_k \tau} = 0 \qquad (3.12)$$

for $i, k = 1, \ldots, 2\mathcal{N}+\mathcal{M}$. Define the functions Φ_{ij}, $i, j = 1, \ldots, 2\mathcal{N}+\mathcal{M}$ as follows:

$$\Phi_{ij} = L_{v_i v_j}(z(\tau), v(\tau), \tau), \qquad i, j = 1, \ldots, 2\mathcal{N}+\mathcal{M}. \qquad (3.13)$$

From (3.13) in (3.12) we obtain the following set of coupled partial differential equations,

$$2\left[\sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ \left(\Phi_{ik}\right)_{z_j} v_j + \left(\Phi_{ik}\right)_{v_j} \Theta_j \right\} + \left(\Phi_{ik}\right)_\tau \right]$$
$$+ \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ \Phi_{ij} \left(\Theta_j\right)_{v_k} + \Phi_{kj} \left(\Theta_j\right)_{v_i} \right\} = 0 \qquad (3.14)$$

for $i, k = 1, \ldots, 2\mathcal{N}+\mathcal{M}$, along a given trajectory segment $z(\tau), v(\tau), \tau \in I$, I a finite interval of R, satisfying (3.9). The total derivative $\frac{d\Phi_{ik}}{d\tau}$ is given by

$$\frac{d\Phi_{ik}}{d\tau} = \left(\Phi_{ik}\right)_\tau + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ \left(\Phi_{ik}\right)_{z_j} v_j + \left(\Phi_{ik}\right)_{v_j} \Theta_j \right\} \qquad (3.15)$$

for $i, k = 1, \ldots, 2\mathcal{N}+\mathcal{M}$. Thus, along the given trajectory, using (3.15) to substitute into (3.14) we obtain

$$2\frac{d\Phi_{ik}}{d\tau} + \sum_{j=1}^{2\mathcal{N}+\mathcal{M}} \left\{ \Phi_{ij} \left(\Theta_j\right)_{v_k} + \Phi_{kj} \left(\Theta_j\right)_{v_i} \right\} = 0$$

for $i, k = 1, \ldots, 2\mathcal{N}+\mathcal{M}$, along a trajectory generated by

$$\frac{dv_l(\tau)}{d\tau} = \Theta_l(z(\tau), v(\tau)), \text{ and } \frac{dz_l(\tau)}{d\tau} = v_l(\tau),$$

for $l = 1, \ldots, 2\mathcal{N}+\mathcal{M}$. □

The proposition provides us with a mechanism to solve the original differential equations in (2.3) - (2.5) by constructing a variational problem of the form (P3). The original system is converted into a system of second order differential equations (3.4), where Θ is determined in equation (3.5). By proposition 3.1, we then solve equation (3.8) for Φ_{ik}. To complete the process, we use Φ_{ik} to construct the Lagrangian. The following corollary provides an expression for a specific type of Lagrangian in terms of Φ_{ik}. Finally, we solve the variational problem (P3) with the Lagrangian given in (3.16) using direct methods [4, 12].

COROLLARY 3.2. *Given Φ_{ik}, for $i, k = 1, \ldots, 2\mathcal{N}+\mathcal{M}$, we can construct a square of a Finsler Lagrangian of the form,*

$$L(z(\tau), v(\tau), \tau) = \sum_{i=1}^{2\mathcal{N}+\mathcal{M}} \sum_{k=1}^{2\mathcal{N}+\mathcal{M}} v_i \Phi_{ik} v_k \qquad (3.16)$$

where $v_i = \frac{dz_i(\tau)}{d\tau}$.

The proof of the corollary is an algebraic exercise, based on Euler's theorem for homogeneous Lagrangians (see [12]).

4. Example. Consider solving the original problem (P1) using barrier functions as a basis for interior point methods [7]. The first step is to reformulate the constrained problem in (P1) as an unconstrained problem with the use of barrier functions. It can be shown [6] that (P1) can be approximated by an unconstrained problem using a barrier function b and barrier coefficients μ_i as,

$$\underset{\xi,x,s}{\text{minimize}}\ F(\xi,x,s,\mu) \tag{ExP1}$$

where $F(\xi,x,s,\mu) = \xi + \sum_{i=0}^{\mathcal{M}_1+3\mathcal{M}_2} \mu_i b\left(g_i\left(\xi,x,s\right)\right)$, ξ is an upper bound on the objective $\tilde{C}(x)$, the constraint functions $g_i(\xi,x,s)$ are,

$$g_i(\xi,x,s) = \begin{cases} \tilde{g}_i\left(x\right) & i = 0,\ldots,\mathcal{M}_1-1 \\ \tilde{h}_{i-\mathcal{M}_1}\left(x\right) - s_{i-\mathcal{M}_1} & i = \mathcal{M}_1,\ldots,\mathcal{M}_1+\mathcal{M}_2-1 \\ s_{i-\mathcal{M}_1-\mathcal{M}_2} - \tilde{h}_{i-\mathcal{M}_1-\mathcal{M}_2}\left(x\right) & i = \mathcal{M}_1+\mathcal{M}_2,\ldots,\mathcal{M}_1+2\mathcal{M}_2-1 \\ s_{i-\mathcal{M}_1-2\mathcal{M}_2} & i = \mathcal{M}_1+2\mathcal{M}_2,\ldots,\mathcal{M}_1+3\mathcal{M}_2-1 \\ \xi - \tilde{C}(x) - \sum_{j=0}^{\mathcal{M}_2-1} r_j s_j & i = \mathcal{M}_1+3\mathcal{M}_2 \end{cases},$$

$x = (x_0, x_1, ..., x_{\mathcal{N}-1})^T \in \mathcal{R}^{\mathcal{N}}$, and $s = (s_0, s_1, ..., s_{\mathcal{M}_2-1})^T \in \mathcal{R}^{\mathcal{M}_2}$. Notice the first $\mathcal{M}_1$ equations of $g_i(\xi,x,s)$ are the original $\mathcal{M}_1$ inequality constraints. The next set of $2\mathcal{M}_2$ equations convert the original $\mathcal{M}_2$ equality constraints into "greater than" and "less than" inequalities and then include additional slack/surplus variables. The following set of $\mathcal{M}_2$ equations allow for non-negativity of the slack/surplus variables, and the final equation is the original objective function with a penalty for the slack/surplus variables. We also modify the notation, letting $\tilde{y} = (\xi, x^T, s^T)^T$, so that the unconstrained problem, for given μ, can be written simply as,

$$\underset{\tilde{y}}{\text{minimize}}\ F(\tilde{y},\mu).$$

It is commonly assumed that a general barrier function b takes on an infinite value for infeasible points, and has a zero value inside the feasible region defined by the constraints of (P1). This is typically implemented with an extended logarithmic barrier function [6], which grows very large as it approaches zero from the right. Hence the choice of this barrier function forces feasibility by driving the constraints to be greater than zero.

An interior point algorithm minimizes $F(\tilde{y},\mu^{(k)})$ for a sequence of positive barrier parameters $\{\mu^{(k)}\}$, such that $\lim_{k\to\infty} \mu^{(k)} = 0$ [7]. Usually the choice of the barrier parameters is ad hoc, but using our meta-control procedure, we determine an optimal sequence $\{\mu^{(k)}\}$ to minimize convergence time.

To continue the example of our meta-control procedure, we identify control variables $\tilde{u}_k$ in (P2) with the barrier parameters $\mu^{(k)}$. We also specify the algorithm to solve (ExP1) to be a Newton descent method. This provides the function $\tilde{f}\left(\tilde{y}_k, \tilde{u}_k\right)$, and we get an iteration of the form (1.1),

$$\tilde{y}_{k+1} = \tilde{y}_k + \tilde{f}\left(\tilde{y}_k, \tilde{u}_k\right)$$

where

$$\tilde{f}(\tilde{y}_k, \tilde{u}_k) = -\left(F_{\tilde{y}\tilde{y}}\right)^{-1}\left(F_{\tilde{y}}\right)^T\Big|_{\tilde{y}_k, \tilde{u}_k} \tag{4.1}$$

and $F_{\tilde{y}}$ is the vector of first partial derivatives of F with respect to $\tilde{y}$, $F_{\tilde{y}\tilde{y}}$ is the Hessian, $\tilde{u}_k$ are the barrier parameters, and $\tilde{y}_k = (\xi_k, x_k^T, s_k^T)^T$, for each iteration k. To continualize the problem, we introduce $y(\tau)$, a continuous version of $\tilde{y}_k$ and $u(\tau)$, a continuous version of $\tilde{u}_k$. Notice the dimension of $y(\tau)$ is the same as $\tilde{y}$, which is $n = \mathcal{N} + \mathcal{M}_2 + 1$. The dimension of $u(\tau)$ equals the dimension of $\tilde{u}$ equals $m = \mathcal{M}_1 + 3\mathcal{M}_2 + 1$. The differential equation satisfied by $y(\tau)$, as in (1.2), is given by

$$\frac{dy(\tau)}{d\tau} = f(y(\tau), u(\tau)) \tag{4.2}$$

where $f(y(\tau), u(\tau)) = -\left(F_{yy}\right)^{-1}\left(F_y\right)^T\Big|_{y(\tau), u(\tau)}$ represents a descent field equation [8]. We assume $f(y(\tau), u(\tau))$ is twice jointly differentiable with respect to y and u.

We select the minimum convergence horizon as the criterion for the optimal control of the algorithm. To provide the criterion for (P2), we would like to set $\Psi(x(T)) = T$ to minimize time, however the optimal control problem assumes T is given. To minimize convergence time, we reparameterize the problem and create an auxiliary convergence time t with a fixed interval, $t \in [0, 1]$. We also introduce a new state variable to represent the "clock", $y_{n+1}(t) = \tau$, where $n = \mathcal{N} + \mathcal{M}_2 + 1$, and a new control variable representing the "clock rate", $u_{m+1}(t) = \frac{dy_{n+1}(t)}{dt}$, where $m = \mathcal{M}_1 + 3\mathcal{M}_2 + 1$. The dynamics become,

$$\begin{aligned} \frac{dy(t)}{dt} &= f(y(t), u(t))u_{m+1}(t) \\ \frac{dy_{n+1}(t)}{dt} &= u_{m+1}(t) \end{aligned}$$

for $t \in [0, 1)$. Now we define the criterion for the optimal control problem. First, we set

$$\Psi = y_{n+1}(1) + \frac{1}{2}\left((F_y)\big|_{y(1),u(1)} \; Q \; (F_y)^T\Big|_{y(1),u(1)} \right)$$

to minimize the convergence time and drive the gradient F_y to zero, where Q is a positive definite matrix. We also set

$$\phi = \frac{1}{2}\left(u(t)^T R u(t) + r u_{m+1}^2(t)\right)$$

for some positive definite matrix R and scalar $r > 0$, to force the control parameters to zero in order to satisfy $\lim_{k\to\infty} \mu^{(k)} = 0$.

Our example of the optimal control problem (P2) becomes,

$$\min_{\substack{u(t), u_{m+1}(t) \\ t\in[0,1)}} \int_0^1 \frac{1}{2} \begin{bmatrix} u^T(t) & u_{m+1}(t) \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} u(t) \\ u_{m+1}(t) \end{bmatrix} dt + y_{n+1}(1) + \frac{1}{2}\left((F_y)\big|_{y(1),u(1)} \; Q \; (F_y)^T\Big|_{y(1),u(1)} \right) \tag{ExP2}$$

subject to

$$\begin{aligned} \frac{dy(t)}{dt} &= f(y(t), u(t))u_{m+1}(t) && \text{for } t \in [0, 1) \\ \frac{dy_{n+1}(t)}{dt} &= u_{m+1}(t) && \text{for } t \in [0, 1) \end{aligned}$$

with initial conditions $y(0) = y_0$, and $y_{n+1}(0) = 0$ given. The Hamiltonian associated with problem (ExP2) is

$$\begin{aligned}\mathcal{H} \quad & (y(t), y_{n+1}(t), u(t), u_{m+1}(t), p(t), p_{n+1}(t)) \\ &= \frac{1}{2} \begin{bmatrix} u^T(t) & u_{m+1}(t) \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} u(t) \\ u_{m+1}(t) \end{bmatrix} \\ &+ \begin{bmatrix} p^T(t) & p_{n+1}(t) \end{bmatrix} \begin{bmatrix} f(y(t), u(t))u_{m+1}(t) \\ u_{m+1}(t) \end{bmatrix}.\end{aligned}$$

Now we present the differential equations for our example, as in (2.3) - (2.5). For $t \in [0, 1)$,the expressions for (2.3) are,

$$\frac{dy(t)}{dt} = f(y(t), u(t))u_{m+1}(t) \tag{4.3}$$

$$\frac{dy_{n+1}(t)}{dt} = u_{m+1}(t) \tag{4.4}$$

with boundary conditions $y\,(0) = y_0$ and $y_{n+1}\,(0) = 0$.

We now construct $\frac{du(t)}{dt} = W$ from (2.4) for our example. As in section 2, the problem of minimizing the Hamiltonian with respect to u, as in the necessary condition (2.1), is similar to the problem of minimizing F with respect to y. Therefore the corresponding descent field for finding optimal trajectories for the continualized barrier parameters is given by

$$\begin{aligned}\frac{d}{dt} \begin{bmatrix} u(t) \\ u_{m+1}(t) \end{bmatrix} &= W(y(t), y_{n+1}(t), u(t), u_{m+1}(t), p(t), p_{n+1}(t)) \\ &= - \left. \frac{\begin{bmatrix} \mathcal{H}_{uu} & \mathcal{H}_{uu_{m+1}} \\ \mathcal{H}^T_{uu_{m+1}} & \mathcal{H}_{u_{m+1}u_{m+1}} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{H}^T_u \\ \mathcal{H}_{u_{m+1}} \end{bmatrix}}{\begin{bmatrix} \mathcal{H}_u & \mathcal{H}_{u_{m+1}} \end{bmatrix} \begin{bmatrix} \mathcal{H}_{uu} & \mathcal{H}_{uu_{m+1}} \\ \mathcal{H}^T_{uu_{m+1}} & \mathcal{H}_{u_{m+1}u_{m+1}} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{H}^T_u \\ \mathcal{H}_{u_{m+1}} \end{bmatrix}} \right|_{\substack{y(t), y_{n+1}(t), \\ u(t), u_{m+1}(t), \\ p(t), p_{n+1}(t)}}\end{aligned} \tag{4.5}$$

with boundary conditions $u\,(0) = u^0$ and $u_{m+1}\,(0) = u^0_{m+1}$.

Finally, we provide the expressions for (2.5) for our example,

$$\frac{dp(t)}{dt} = -u_{m+1}(t) \left(\frac{\partial f(y(t), u(t))}{\partial y} \right)^T p(t) \tag{4.6}$$

$$\frac{dp_{n+1}(t)}{dt} = 0 \tag{4.7}$$

with boundary conditions

$$p(1) = \left. (F_{yy}) \right|_{y(1),u(1)} \; Q \; \left. (F_y)^T \right|_{y(1),u(1)} \tag{4.8}$$

$$p_{n+1}(1) = 1 \tag{4.9}$$

given. Notice that the derivative of p_{n+1} is identically zero, which implies that p_{n+1} is a constant equal to one on the interval $[0, 1]$. This is a consequence of p_{n+1} being the costate of the "clock". Equations (4.3) - (4.9) constitute the necessary conditions for optimality of problem (ExP2). In order to solve this system of differential equations, we use the inverse Lagrangian method described in section 3.

Following the inverse Lagrangian procedure in section 3, we first introduce a vector-valued function $z : [0,1) \to \mathcal{R}^{2(n+1)+(m+1)}$ as follows

$$z(t) = \begin{bmatrix} y(t) \\ y_{n+1}(t) \\ u(t) \\ u_{m+1}(t) \\ p(t) \\ p_{n+1}(t) \end{bmatrix}, \quad t \in [0,1).$$

Then, we substitute z into the system of differential equations (4.3) - (4.9) and take the derivative with respect to t, as in equation (3.4). The resulting vector function Θ for our example is,

$$\Theta\left(z(t), \frac{dz(t)}{dt}, t\right) = \begin{bmatrix} u_{m+1}(t)\left(f_y\left(y(t),u(t)\right)\frac{dy(t)}{dt} + f_u\left(y(t),u(t)\right)\frac{du(t)}{dt}\right) \\ \quad +f\left(y(t),u(t)\right)\frac{du_{m+1}(t)}{dt} \\ \frac{du_{m+1}(t)}{dt} \\ W_y\frac{dy(t)}{dt} + W_{y_{n+1}}\frac{dy_{n+1}(t)}{dt} + W_u\frac{du(t)}{dt} + W_{u_{m+1}}\frac{du_{m+1}(t)}{dt} + \\ \quad +W_p\frac{dp(t)}{dt} + W_{p_{n+1}}\frac{dp_{n+1}(t)}{dt} \\ -u_{m+1}(t)\left(f_{yy}\left(y(t),u(t)\right)\frac{dy(t)}{dt} + f_{yu}\left(y(t),u(t)\right)\frac{du(t)}{dt}\right)^T p(t) \\ \quad -u_{m+1}(t)f_y^T\left(y(t),u(t)\right)\frac{dp(t)}{dt} - f_y^T\left(y(t),u(t)\right)p(t)\frac{du_{m+1}(t)}{dt} \end{bmatrix}$$

Knowing $\Theta\left(z(t), \frac{dz(t)}{dt}, t\right)$ allows us to determine Φ_{ik}, for $i,k = 1,\ldots,2(n+1)+(m+1)$, as in proposition 3.1, and then determine the Lagrangian

$$L(z(t), v(t), t) = \sum_{i=1}^{2(n+1)+(m+1)} \sum_{k=1}^{2(n+1)+(m+1)} v_i \Phi_{ik} v_k$$

using corollary 3.2. This gives us a variational problem for (P3),

$$\min_{\substack{z(t)\\ t\in[0,1)}} \int_0^1 L\left(z(t), \frac{dz(t)}{dt}, t\right) dt \tag{ExP3}$$

which can be solved using direct methods. Direct methods include the Ritz method, the method of finite differences [10], and more recently finite element methods [4] and the chattering approximation method [9]. The solution to the variational problem (ExP3), $z(1)$, provides the optimal solution to the original problem (ExP1). Thus, the first entry of $z(1)$, ξ, provides an approximation (upper bound) of the optimal objective function value, the next $\mathcal{N}$ entries of $z(1)$ are the optimal values for the original variables x, and the following $\mathcal{M}_2$ entries of $z(1)$ are the slack/surplus variables for the equality constraints which should be very close to zero. The rest of the values of $z(1)$ are algorithmic parameters, including the values of the barrier parameters. This example demonstrates our meta-control approach that simultaneously optimizes the original problem and barrier parameter values.

5. Conclusion. In this paper, we present an approach to solve an optimization problem while controlling the algorithm simultaneously. We use optimal control theory in order to compute control parameters that steer the algorithm to provide best performance. We first formulate an optimal control problem, and construct a set of differential equations from the necessary conditions of optimality. We then use the inverse Lagrangian method to construct a corresponding variational problem, which can be solved with direct methods. The solutions to the variational problem provide both the optimal algorithmic parameters as well as the best path to the optimal solution of the original problem. We demonstrate our approach using an example to control a Newton-like iteration for solving a constrained nonlinear optimization problem using interior point method. We measure performance of the algorithm in terms of convergence time, and the algorithmic parameters are the barrier coefficients.

REFERENCES

[1] M. Athans and P. L. Falb. *Optimal Control*. McGraw-Hill Book Company, New York, N.Y., 1966.

[2] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley, Hoboken, New Jersey, second edition, 2003.

[3] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1996.

[4] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. *Computational Differential Equations*. Cambridge University Press, Lund, Sweden, 1996.

[5] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, Rhode Island, 1998.

[6] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1990.

[7] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44(4):525–597, 2002.

[8] Theodore Frankel. *The Geometry of Physics*. Cambridge University Press, New York, 1997.

[9] Xiaolin Ge, Wolf Kohn, Anil Nerode, and Jeffrey B. Remmel. Hybrid systems: chattering approximations to relaxed controls. In R. Alur, T.A. Henzinger, and E.D. Sontag, editors, *Hybrid Systems III*, number 1066 in Lecture Notes in Computer Science, page 76. Springer.

[10] I. M. Gelfand and S. V. Fomin. *Calculus of variations*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[11] Wolf Kohn and Vladimir Brayman. Convergence analysis of the hynomics incremental optimizer. Technical report, Hynomics Corporation, 2003.

[12] Wolf Kohn, Vladimir Brayman, Pawel Cholewinski, and Anil Nerode. Control in hybrid systems. *International Journal of Hybrid Systems*, 2004.

[13] Wolf Kohn, Anil Nerode, and Jeffrey B. Remmel. Continualization: A hybrid systems control technique for computing. In *Proceedings of CESA'96 IMACS Multiconference*, pages 517–521, 1996.

[14] Harold J. Kushner and Dean S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, New York, N.Y., 1978.

[15] Harold J. Kushner and G. George Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York, N.Y., 1997.

[16] G. P. McCormick. *Nonlinear Programming: Theory, Algorithms, and Applications*. John Wiley and Sons, New York, 1983.

[17] Peter J. Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag, second edition, 1993.

[18] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gankrelidze, and E. F. Mischenko. *The Mathematical Theory of Optimal Processes*. Interscience, New York, 1962.