2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Locally Optimal and Globally Inverse Optimal Controller for Multi-Input Nonlinear Systems

Hongkeun Kim, Juhoon Back, Hyungbo Shim, and Jin Heon Seo

***Abstract*— The $\mathcal{H}_\infty$-optimal control for nonlinear systems is hard to obtain because one must solve the Hamilton-Jacobi-Isaacs (HJI) equation. To overcome this problem, a nonlinear controller is proposed by Ezal, Pan, and Kokotović. The controller guarantees local optimality and global inverse optimality, that is, it behaves as a linear optimal controller in the region where the linearized dynamics dominates, and is inverse optimal in the global sense. However, the system class under their consideration is *single-input* strict-feedback nonlinear systems which is somewhat restrictive. In this paper, we propose a nonlinear optimal controller for a class of *multi-input* nonlinear systems. Moreover, under the proposed controller, the closed-loop system is globally exponentially stable, whereas the controller proposed by Ezal *et al.* just guarantees global asymptotic stability.**

I. Introduction

In general, the nonlinear optimal control designs, including the $\mathcal{H}_\infty$-design, for the general nonlinear systems are very difficult since one must solve the HJI equation (see [2] and [4]). An alternative to this problem is to employ the linear optimal control which is obtained from the linearization of the given nonlinear system. However, the main drawback of this approach is global stability, that is, the closed-loop system constituted by the linear optimal controller is locally stable but may be unstable in the outside of the local region. Moreover, one does not know how large the region of attraction is. It may become unacceptably small in the general cases. To overcome this problem, Ezal *et al.* [7] proposed a nonlinear controller based on the robust backstepping methodology, whose feature is that the closed-loop system is locally optimal and globally asymptotically stable. In particular, the global stability is achieved by ensuring the inverse optimality so that the closed-loop system has desirable stability margins [2].

However, the considered systems in [7] are single-input strict-feedback nonlinear systems, which might be somewhat restrictive. The extension to the multi-input nonlinear systems is presented in [8]. But, in [8], the authors only consider a specific nonlinear model, i.e., a moored ship model. Due to the consideration of the specific nonlinear model, they only perform the backstepping procedure to two steps. Furthermore, the dimensions of the substates must be equal to each other.

The main objective of this paper is to extend the approach developed in [7] to the general *multi-input* nonlinear systems. The essential difference compared to [7] is the *block Cholesky factorization*. By using the *block Cholesky factorization* and the *block* backstepping methodology, a nonlinear controller is constructed such that it guarantees the local optimality and the global inverse optimality for the *multi-input* nonlinear systems.

H. Kim, H. Shim, and J.H. Seo are with ASRI, School of Electrical Engineering and Computer Science, Seoul National University, Seoul, Korea (e-mail: hkkim@cdsl.kr).

J. Back is with Department of Mechanical Engineering, Korea University, Seoul, Korea.

II. Problem Formulation

Consider a *multi-input* nonlinear system

$$\dot{x} = f(x) + G(x)u + H(x)w,$$

in the *block* strict-feedback form [3]:

$$\begin{aligned} \dot{x}_1 &= f_1(x_{[1]}) + G_1(x_{[1]})x_2 + H_1(x_{[1]})w \\ \dot{x}_2 &= f_2(x_{[2]}) + G_2(x_{[2]})x_3 + H_2(x_{[2]})w \\ &\vdots \\ \dot{x}_m &= f_m(x) + G_m(x)u + H_m(x)w, \end{aligned} \tag{1}$$

where $x = [x_1^T \ \cdots \ x_m^T]^T$ with $x_i \in \mathbb{R}^{\mu_i}$ is the state, $x_{[i]} = [x_1^T \ \cdots \ x_i^T]^T$ is the substate of x, $u \in \mathbb{R}^{\mu_{m+1}}$ is the control input, the positive integer μ_i satisfies $\mu_1 + \cdots + \mu_m = n$ with $0 < \mu_1 \leq \cdots \leq \mu_m \leq \mu_{m+1}$, and $w : [0, \infty) \to \mathbb{R}^q$ is an unknown disturbance of either $\mathcal{L}_2$ or $\mathcal{L}_\infty$. We assume that f_i, G_i, H_i are smooth, $f_i(0) = 0$, and $G_i(x_{[i]})$ has full row rank for all $x_{[i]} \in \mathbb{R}^{\mu_1 + \cdots + \mu_i}$. Note that the system (1) is a natural extension of the system in [7], that is, if $m = n$ and $G_i(x_{[i]}) = 1$ for all i then (1) becomes the system considered in [7].

The goal of this paper is to develop a recursive design procedure for the *multi-input* nonlinear system (1) such that the designed controller $u = \mu(x)$ achieves *local optimality* and *global inverse optimality* as described below.

Local Optimality: We design the controller $u = \mu(x)$ to satisfy the local optimality, that is, to make it similar to the $\mathcal{H}_\infty$-optimal controller in the region where the linear dynamics dominates. Consider the linear part of (1)

$$\dot{x}_l = Ax_l + Bu_l + Dw_l, \tag{2}$$

where subscript "l" denotes local quantities and

$$A = \frac{\partial f}{\partial x}(0) = \begin{bmatrix} A_{11} & B_1 & 0 & \cdots & 0 \\ A_{21} & A_{22} & B_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & B_{m-1} \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mm} \end{bmatrix}, \tag{3}$$

$$B = G(0) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ B_m \end{bmatrix}, \quad D = H(0) = \begin{bmatrix} D_1 \\ \vdots \\ D_{m-1} \\ D_m \end{bmatrix}.$$

Since B_i's have full row rank, the pair (A, B) is controllable. Suppose that a cost functional to (2) is given by

$$J_l(u_l, w_l) = \int_0^\infty \left\{ x_l^T Q_l x_l + u_l^T R_l u_l - \gamma^2 w_l^T w_l \right\} dt, \quad (4)$$

with $Q_l = Q_l^T > 0$ and $R_l = R_l^T > 0$. Then, the $\mathcal{H}_\infty$-optimal controller $u_l = \mu_l(x_l)$ is the solution of the *dynamic game* $\min_{u_l} \max_{w_l} J_l(u_l, w_l)$ for the linearized system (2) and the cost functional (4). From the $\mathcal{H}_\infty$-optimal control theory, it is known that there exist the optimal disturbance attenuation level $\gamma^* > 0$ and the unique positive definite solution P of the *generalized algebraic Riccati equation* (GARE)

$$PA + A^T P + P \left(\frac{1}{\gamma^2} DD^T - BR_l^{-1} B^T \right) P + Q_l = 0 \quad (5)$$

for all $\gamma > \gamma^*$, because (A, B) is controllable and (A, Q_l) is observable [4]. In this case, the optimal value function is $V(x_l) = x_l^T P x_l$ (see [2] and [5]), the resulting optimal controller for the cost functional (4) is

$$u_l = \mu_l(x_l) = -R_l^{-1} B^T P x_l, \quad (6)$$

and the corresponding worst case disturbance is

$$w_l = \nu_l(x_l) = \frac{1}{\gamma^2} D^T P x_l. \quad (7)$$

Thus, in order to satisfy the local optimality, we require that the final controller $u = \mu(x)$ satisfies $\frac{\partial \mu}{\partial x}(0)x = \mu_l(x)$.

Global Inverse Optimality: We construct a controller u that achieves the global inverse optimality [2] for the *multi-input* nonlinear system (1) with respect to the cost functional

$$J(u, w) = \int_0^\infty \left\{ q(x) + u^T R(x) u - \gamma^2 w^T w \right\} dt, \quad (8)$$

where $q(x)$ is positive definite and radially unbounded, and $R(x) = R^T(x) > 0$ for all x. Note that, differently from the one in [7] that is a scalar-valued positive function, $R(x)$ should be symmetric positive definite for all x because *multi-input* nonlinear systems are considered. Moreover, $q(x)$ and $R(x)$ satisfy that $q_{xx}(0) := \frac{1}{2} \frac{\partial^2 q}{\partial x^2}(0) = Q_l$ and $R(0) = R_l$ for the local optimality to be meaningful. The scaling factor $1/2$ is included for convenience. The inverse optimality problem, which is mentioned above, is equivalent to satisfying

$$\min_u \max_w \left\{ q(x) + u^T R(x) u - \gamma^2 w^T w + \dot{V}(x) \right\} = 0. \quad (9)$$

for some value function $V : \mathbb{R}^n \to \mathbb{R}$.

III. Locally Optimal and Globally Inverse Optimal Controller Design

Our objective is to design a stabilizing nonlinear state-feedback controller, which is globally inverse optimal and whose linear part is the same as the linear optimal controller (6). The resulting controller is obtained by the robust back-stepping, but in order to have the latter property, the linear part of each virtual control that is obtained at each step of the backstepping design procedure, needs to be carefully chosen. For this, the linear optimal controller (6) is obtained first. Then, the *block Cholesky factorization* of the matrix P gives some indication suitable for the linear parts of all the virtual controls. We, therefore, finalize the controller design for the nonlinear system (1) by choosing nonlinear virtual controls at each step using the linear parts that are already known to us. The obtained controller is shown to be optimal for a certain cost functional.

Before proceeding, we introduce some notations for convenience. Consider a matrix $M \in \mathbb{R}^{n \times n}$ which is partitioned into the same form as A in (3), that is, its (i, j) matrix entry M_{ij} belongs to $\mathbb{R}^{\mu_i \times \mu_j}$. Then, it is defined that

$$M_{[i]} := \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1i} \\ \vdots & \vdots & \ddots & \vdots \\ M_{i1} & M_{i2} & \cdots & M_{ii} \end{bmatrix}$$

and $M_{\langle i \rangle} := [M_{i1} \ M_{i2} \ \cdots \ M_{ii}]$. For a matrix $N = [N_1^T \ \cdots \ N_m^T]^T$ with $N_i \in \mathbb{R}^{\mu_i \times q}$, it is defined that $N_{[i]} := [N_1^T \ \cdots \ N_i^T]^T$. To represent a vector we use a small letter and to denote a matrix a capital letter is employed. The superscript "h" stands for higher order function. In this paper, the higher order function has two meanings. For a vector function $a(x)$, $a^h(x)$ means that $a^h(0) = 0$ and $\frac{\partial a^h}{\partial x}(0) = 0$. For a matrix function $C(x)$, $C^h(x)$ is intended for $C^h(0) = 0$. The right inverse is denoted by "+", i.e., $K^+ := K^T(KK^T)^{-1}$ for full row rank matrix K. With a certain coordinate transformation, the "over bar" means a quantity in the transformed coordinate. Finally the norm operator $\|\cdot\|$ is devoted for the Euclidean norm.

A. Preliminaries

As shown in [7], it is necessary to construct (6) and (7) by using the linear backstepping methodology for the preparation of nonlinear backstepping. Instead of constructing them explicitly, the forthcoming Lemma 1 will summarize the necessary properties. Consider the unique positive definite solution P of the GARE (5). From Lemma 3 in Appendix, P can be factorized as $P = L^T \Delta L$, where

$$L = \begin{bmatrix} I_1 & 0 & \cdots & 0 \\ L_{11} & I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{m-1,1} & L_{m-1,2} & \cdots & I_m \end{bmatrix},$$
$$\Delta = \mathrm{diag}(\Delta_1, \Delta_2, \cdots, \Delta_m)$$

with $\Delta_i, \ I_i \in \mathbb{R}^{\mu_i \times \mu_i}$, $\Delta_i = \Delta_i^T > 0$ for $i = 1, \cdots, m$, and other matrices are of appropriate dimensions. Define $z_l :=$

Lx_l. In particular, the error coordinates are given by

$$\begin{aligned} z_{l,1} &= x_{l,1} \\ z_{l,i} &= x_{l,i} + L_{\langle i-1\rangle} x_{l,[i-1]} \\ &= x_{l,i} + \overline{L}_{\langle i-1\rangle} z_{l,[i-1]}, \; i = 2, \cdots, m, \end{aligned} \tag{10}$$

where $\overline{L}_{\langle i\rangle} := L_{\langle i\rangle} L_{[i]}^{-1}$. Then, the dynamics of z becomes

$$\dot{z}_l = \overline{A} z_l + \overline{B} u_l + \overline{D} w_l, \tag{11}$$

where $\overline{A} = LAL^{-1}$, $\overline{B} = LB$, and $\overline{D} = LD$.

Lemma 1: The following properties hold.

P1: For $1 \le k \le m$, $z_{l,[k]} = L_{[k]} x_{l,[k]}$, where $L_{[k]}$ is invertible and $L_{[k]}^{-1} := (L^{-1})_{[k]} = (L_{[k]})^{-1}$ has the same structure as $L_{[k]}$ (see M1 in Appendix).

P2: The structure of $\overline{A}$ is the same as that of A, that is,

$$\overline{A} = \begin{bmatrix} \overline{A}_{11} & B_1 & 0 & \cdots & 0 \\ \overline{A}_{21} & \overline{A}_{22} & B_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & B_{m-1} \\ \overline{A}_{m1} & \overline{A}_{m2} & \overline{A}_{m3} & \cdots & \overline{A}_{mm} \end{bmatrix},$$

and $\overline{B} = B$. Thus, for $1 \le k < m$, the dynamics of the substate $z_{l,[k]}$ can be written as

$$\dot{z}_{l,[k]} = \overline{A}_{[k]} z_{l,[k]} + \begin{bmatrix} 0_{k-1} \\ B_k z_{l,k+1} \end{bmatrix} + \overline{D}_{[k]} w_l,$$

where 0_k is $(\mu_1 + \cdots + \mu_k) \times 1$ zero vector.

P3: In the z-coordinate, the GARE (5) becomes

$$\Delta \overline{A} + \overline{A}^T \Delta + \Delta \left(\frac{1}{\gamma^2} \overline{D}\,\overline{D}^T - B R_l^{-1} B^T \right) \Delta + \overline{Q} = 0,$$

where $\overline{Q} = L^{-T} Q_l L^{-1} > 0$. In particular, for $1 \le k < m$,

$$\begin{aligned} \Delta_{[k]} \overline{A}_{[k]} + \overline{A}_{[k]}^T \Delta_{[k]} + \frac{1}{\gamma^2} \Delta_{[k]} \overline{D}_{[k]} \overline{D}_{[k]}^T \Delta_{[k]} + \overline{Q}_{[k]} = 0, \\ 2 z_{l,[k]}^T \Delta_{[k]} \overline{A}_{[k]} z_{l,[k]} = - z_{l,[k]}^T \overline{Q}_{[k]} z_{l,[k]} - \gamma^2 \overline{\nu}_{lk}^T \overline{\nu}_{lk}, \end{aligned}$$

where $\overline{\nu}_{lk}(z_{l,[k]}) := \frac{1}{\gamma^2} \overline{D}_{[k]}^T \Delta_{[k]} z_{l,[k]}$.

P4: For $1 \le k \le m$,

$$\begin{aligned} z_{l,[k]}^T \Delta_{[k]} \overline{A}_{[k]} z_{l,[k]} &= z_{l,[k-1]}^T \Delta_{[k-1]} \overline{A}_{[k-1]} z_{l,[k-1]} \\ &\quad + z_{l,k-1}^T \Delta_{k-1} B_{k-1} z_{l,k} + z_{l,k}^T \Delta_k \overline{A}_{\langle k\rangle} z_{l,[k]}. \end{aligned}$$

P5: For $1 \le k \le m$,

$$\overline{\nu}_{lk}(z_{l,[k]}) = \overline{\nu}_{l,k-1}(z_{l,[k-1]}) + \frac{1}{\gamma^2} \overline{D}_k^T \Delta_k z_{l,k},$$

with $\overline{\nu}_{l0} \equiv 0$ and $\overline{\nu}_l := \overline{\nu}_{lm}$.

P6: For $1 \le k \le m$,

$$\begin{aligned} \overline{A}_{\langle k\rangle} &= A_{\langle k\rangle} L_{[k]}^{-1} - B_k \overline{L}_{\langle k\rangle} + \left[\overline{L}_{\langle k-1\rangle} \overline{A}_{[k-1]} \quad \overline{L}_{k-1,k-1} B_{k-1} \right] \\ \overline{D}_k &= D_k + \overline{L}_{\langle k-1\rangle} \overline{D}_{[k-1]}, \end{aligned}$$

where $L_{\langle 0\rangle} = 0$ and $L_{\langle m\rangle} = 0$.

Lemma 1 lists the properties in [7] for *multi-input* case and its proof is omitted since it is similar to that of [7]. With the help of Lemma 1 and the linear backstepping procedure in [7], the control (6) and disturbance (7) can be recovered, i.e., in z-coordinate, $u_l = \overline{\mu}_l(z_l) = -R_l^{-1} B^T \Delta z_l$ and $w_l = \overline{\nu}_l(z_l) = \frac{1}{\gamma^2} \overline{D}^T \Delta z_l$.

B. Nonlinear Backstepping

Based on the properties in Lemma 1, we proceed the nonlinear backstepping. At each step of nonlinear backstepping, the new state $z_i = \phi_i(x_{[i]})$ will be constructed such that its linear part is the same as the previous one, i.e., (10). The only difference of the new state z_i compared to the linear one is higher order terms in the virtual control to cancel the nonlinearities and to attenuate the disturbance. At the last step, the control Lyapunov function (CLF) [6] $\overline{V}(z) = z^T \Delta z$ and the lower triangular diffeomorphism $z = \phi(x)$, where $z_i = \phi_i(x_{[i]})$, are obtained. For convenience of the nonlinear backstepping, we define the error coordinates, in advance, as

$$\begin{aligned} z_1 &= \phi_1(x_{[1]}) = x_1 \\ z_i &= \phi_i(x_{[i]}) = x_i + \overline{\alpha}_{i-1}(z_{[i-1]}), \quad i = 2, \cdots, m, \end{aligned} \tag{12}$$

where $\overline{\alpha}_i(z_{[i]}) := \overline{L}_{\langle i\rangle} z_{[i]} + \overline{\alpha}_i^h(z_{[i]})$. Here, $\overline{L}_{\langle i\rangle} z_{[i]}$ is the linear virtual control as in (10) and $\overline{\alpha}_i^h(z_{[i]})$ is a higher order virtual control to be designed at each step.

Step 1: Let $\overline{V}_1(z_{[1]}) = z_{[1]}^T \Delta_{[1]} z_{[1]}$ as the value function. Define $f_1^h(z_{[1]}) := f_1(x_{[1]}) - A_{\langle 1\rangle} x_{[1]}$ and $G_1^h(z_{[1]}) := G_1(x_{[1]}) - B_1$. Note that $f_1^h(0) = 0$, $\frac{\partial f_1^h}{\partial z_{[1]}}(0) = 0$, and $G_1^h(0) = 0$. Then, the dynamics of $z_{[1]}$-subsystem is

$$\begin{aligned} \dot{z}_1 &= \left(A_{\langle 1\rangle} - B_1 \overline{L}_{\langle 1\rangle} \right) z_{[1]} + G_1(z_{[1]}) z_2 + H_1(z_{[1]}) w \\ &\quad + f_1^h(z_{[1]}) - G_1^h(z_{[1]}) \overline{L}_{\langle 1\rangle} z_{[1]} - G_1(z_{[1]}) \overline{\alpha}_1^h(z_{[1]}) \\ &= \overline{A}_{\langle 1\rangle} z_{[1]} + \overline{G}_1(z_{[1]}) z_2 + \overline{H}_1(z_{[1]}) w + \overline{f}_1^h(z_{[1]}), \end{aligned}$$

where $\overline{G}_1(z_{[1]}) := G_1(z_{[1]})$, $\overline{H}_1(z_{[1]}) := H_1(z_{[1]})$, $\overline{f}_1^h(z_{[1]}) := f_1^h(z_{[1]}) - G_1^h(z_{[1]}) \overline{L}_{\langle 1\rangle} z_{[1]} - G_1(z_{[1]}) \overline{\alpha}_1^h(z_{[1]})$, and the error coordinate z_2 is defined in (12). It is easily seen that $\overline{f}_1^h$ is indeed higher order function if $\overline{\alpha}_1^h$ is. To obtain the above representation, P6 in Lemma 1 is employed for $k = 1$. The time derivative of the value function $\overline{V}_1$ becomes

$$\begin{aligned} \dot{\overline{V}}_1 &= 2 z_{[1]}^T \Delta_{[1]} \left\{ \overline{A}_{\langle 1\rangle} z_{[1]} + \overline{G}_1 z_2 + \overline{H}_1 w + \overline{f}_1^h \right\} \\ &= - z_{[1]}^T \overline{Q}_{[1]} z_{[1]} - \gamma^2 \overline{\nu}_{l1}^T \overline{\nu}_{l1} + 2 z_1^T \Delta_1 \overline{G}_1 z_2 + 2\gamma^2 \overline{\nu}_1 w \\ &\quad + 2 z_1^T \Delta_1 \overline{f}_1^h + \gamma^2 w^T w - \gamma^2 w^T w + \gamma^2 \overline{\nu}_1^T \overline{\nu}_1 - \gamma^2 \overline{\nu}_1^T \overline{\nu}_1, \end{aligned}$$

where $\overline{\nu}_1(z_{[1]}) := \frac{1}{\gamma^2} \overline{H}_{[1]}^T (z_{[1]}) \Delta_{[1]} z_{[1]}$. Here, P3 in Lemma 1 is used for $k = 1$. Completing the square with respect to w, we obtain

$$\begin{aligned} \dot{\overline{V}}_1 &= - z_{[1]}^T \overline{Q}_{[1]} z_{[1]} + \gamma^2 w^T w - \gamma^2 \| w - \overline{\nu}_1 \|^2 + 2 z_1^T \Delta_1 \overline{G}_1 z_2 \\ &\quad + 2 z_1^T \Delta_1 \left\{ \overline{f}_1^h + \frac{1}{2\gamma^2} \left(\overline{H}_1 \overline{H}_1^T - \overline{D}_1 \overline{D}_1^T \right) \Delta_1 z_1 \right\}. \end{aligned}$$

To treat the terms in the braces, $\overline{\alpha}_1^h$ is designed as

$$\overline{\alpha}_1^h = G_1^+ \left[f_1^h - G_1^h \overline{L}_{\langle 1\rangle} z_{[1]} + \frac{1}{2\gamma^2} \left(\overline{H}_1 \overline{H}_1^T - \overline{D}_1 \overline{D}_1^T \right) \Delta_1 z_1 \right]$$

and this results in

$$\dot{\overline{V}}_1 = - z_{[1]}^T \overline{Q}_{[1]} z_{[1]} + \gamma^2 w^T w - \gamma^2 \| w - \overline{\nu}_1 \|^2 + 2 z_1^T \Delta_1 \overline{G}_1 z_2.$$

Note that $\overline{\alpha}_1^h$ is made up of only higher order terms, that is, $\overline{\alpha}_1^h(0) = 0$ and $\frac{\partial \overline{\alpha}_1^h}{\partial z_{[1]}}(0) = 0$.

Step i: Select $\overline{V}_i(z_{[i]}) = \overline{V}_{i-1}(z_{[i-1]}) + z_i^T\Delta_i z_i = z_{[i]}^T\Delta_{[i]}z_{[i]}$ as the value function for this step. For the induction, it is assumed that the dynamics of the $z_{[i-1]}$-subsystem is represented by

$$\dot{z}_{[i-1]} = \overline{A}_{[i-1]}z_{[i-1]} + \begin{bmatrix} \overline{g}^h_{[i-2]}(z_{[i-1]}) \\ \overline{G}_{i-1}(z_{[i-1]})z_i \end{bmatrix} + \overline{H}_{[i-1]}(z_{[i-1]})w + \overline{f}^h_{[i-1]}(z_{[i-1]}). \tag{13}$$

Here, we have used the following definitions:

$$z_{[i]} = \phi_{[i]}(x_{[i]}) := [\phi_1^T(x_{[1]}) \ \cdots \ \phi_i^T(x_{[i]})]^T,$$
$$\overline{G}_i(z_{[i]}) := G_i(\phi_{[i]}^{-1}(z_{[i]})), \quad \overline{G}^h_i(z_{[i]}) := \overline{G}_i(z_{[i]}) - B_i,$$
$$\overline{g}^h_{[i-1]}(z_{[i]}) := [\{\overline{G}^h_1(z_{[1]})z_2\}^T \ \cdots \ \{\overline{G}^h_{i-1}(z_{[i-1]})z_i\}^T]^T,$$
$$\overline{f}^h_{[i]}(z_{[i]}) := [\{\overline{f}^h_1(z_{[1]})\}^T \ \cdots \ \{\overline{f}^h_i(z_{[i]})\}^T]^T.$$

Moreover, time derivative of $\overline{V}_{i-1}$ is assumed to be

$$\begin{aligned}\dot{\overline{V}}_{i-1} &= -z_{[i-1]}^T\overline{Q}_{[i-1]}z_{[i-1]} + \gamma^2 w^T w - \gamma^2\|w - \overline{\nu}_{i-1}\|^2 \\ &\quad + 2z_{i-1}^T\Delta_{i-1}\overline{G}_{i-1}z_i \\ &= 2z_{[i-1]}^T\Delta_{[i-1]}\overline{A}_{[i-1]}z_{[i-1]} + \gamma^2\overline{\nu}_{l,i-1}^T\overline{\nu}_{l,i-1} \\ &\quad + 2\gamma^2\overline{\nu}_{i-1}^T w - \gamma^2\overline{\nu}_{i-1}^T\overline{\nu}_{i-1} + 2z_{i-1}^T\Delta_{i-1}\overline{G}_{i-1}z_i,\end{aligned} \tag{14}$$

where $\overline{\nu}_{i-1}(z_{[i-1]}) := \frac{1}{\gamma^2}\overline{H}_{[i-1]}^T(z_{[i-1]})\Delta_{[i-1]}z_{[i-1]}$.

In the followings, it will be shown that the dynamics of the $z_{[i]}$-subsystem can be written as the form in (13) and the time derivative of $\overline{V}_i$ is of the form in (14). To show the former, let $f_i^h(z_{[i]}) := f_i(\phi_{[i]}^{-1}(z_{[i]})) - A_{\langle i\rangle}\phi_{[i]}^{-1}(z_{[i]})$ and $(\phi_{[i]}^{-1})^h(z_{[i]}) := \phi_{[i]}^{-1}(z_{[i]}) - L_{[i]}^{-1}z_{[i]}$. Then, by the definition of z_i in (12) and P6, the dynamics of z_i becomes

$$\begin{aligned}\dot{z}_i &= f_i(x_{[i]}) + G_i(x_{[i]})x_{i+1} + H_i(x_{[i]})w \\ &\quad + \frac{\partial\overline{\alpha}_{i-1}}{\partial z_{[i-1]}}\left\{\overline{A}_{[i-1]}z_{[i-1]} + \begin{bmatrix} \overline{g}^h_{[i-2]}(z_{[i-1]}) \\ \overline{G}_{i-1}(z_{[i-1]})z_i \end{bmatrix}\right. \\ &\quad \left. + \overline{H}_{[i-1]}(z_{[i-1]})w + \overline{f}^h_{[i-1]}(z_{[i-1]})\right\} \\ &= \overline{A}_{\langle i\rangle}z_{[i]} + \overline{G}_i(z_{[i]})z_{i+1} + \overline{H}_i(z_{[i]})w + \overline{f}^h_i(z_{[i]}),\end{aligned}$$

where

$$\begin{aligned}\overline{f}^h_i(z_{[i]}) &:= A_{\langle i\rangle}(\phi_{[i]}^{-1})^h(z_{[i]}) + f_i^h(z_{[i]}) - \overline{G}^h_i(z_{[i]})\overline{L}_{\langle i\rangle}z_{[i]} \\ &\quad + \overline{L}_{\langle i-1\rangle}\overline{g}^h_{[i-1]}(z_{[i]}) \\ &\quad + \frac{\partial\overline{\alpha}^h_{i-1}}{\partial z_{[i-1]}}\left\{\overline{A}_{[i-1]}z_{[i-1]} + \begin{bmatrix} \overline{g}^h_{[i-2]}(z_{[i-1]}) \\ \overline{G}_{i-1}(z_{[i-1]})z_i \end{bmatrix}\right\} \\ &\quad + \frac{\partial\overline{\alpha}_{i-1}}{\partial z_{[i-1]}}\overline{f}^h_{[i-1]}(z_{[i-1]}) - \overline{G}_i(z_{[i]})\overline{\alpha}^h_i(z_{[i]}),\end{aligned}$$
$$\overline{H}_i(z_{[i]}) := H_i(\phi_{[i]}^{-1}(z_{[i]})) + \frac{\partial\overline{\alpha}_{i-1}}{\partial z_{[i-1]}}\overline{H}_{[i-1]}(z_{[i-1]}).$$

Therefore, by using P2, the $z_{[i]}$-dynamics is obtained as

$$\dot{z}_{[i]} = \overline{A}_{[i]}z_{[i]} + \begin{bmatrix} \overline{g}^h_{[i-1]}(z_{[i]}) \\ \overline{G}_i(z_{[i]})z_{i+1} \end{bmatrix} + \overline{H}_{[i]}(z_{[i]})w + \overline{f}^h_{[i]}(z_{[i]}),$$

which shows that the former case is indeed true. Differently from [7], the term $\overline{g}^h_{[i-1]}(z_{[i]})$ appears in $z_{[i]}$-dynamics due to the consideration of the nonlinear function $\overline{G}_{i-1}$. To show the latter, (14), P3, and P4 are used and this results in

$$\begin{aligned}\dot{\overline{V}}_i &= \dot{\overline{V}}_{i-1} + 2z_i^T\Delta_i\left(\overline{A}_{\langle i\rangle}z_{[i]} + \overline{G}_i z_{i+1} + \overline{H}_i w + \overline{f}^h_i\right) \\ &= -z_{[i]}^T\overline{Q}_{[i]}z_{[i]} + \gamma^2 w^T w - \gamma^2\|w - \overline{\nu}_i\|^2 + 2z_i^T\Delta_i\overline{G}_i z_{i+1} \\ &\quad + 2z_i^T\Delta_i\left\{\Delta_i^{-1}(\overline{G}^h_{i-1})^T\Delta_{i-1}z_{i-1} + \overline{f}^h_i\right. \\ &\quad \left. + (\overline{H}_i\overline{\nu}_{i-1} - \overline{D}_i\overline{\nu}_{l,i-1}) + \frac{1}{2\gamma^2}\left(\overline{H}_i\overline{H}_i^T - \overline{D}_i\overline{D}_i^T\right)\Delta_i z_i\right\}\end{aligned}$$

where $\overline{\nu}_i(z_{[i]}) := \overline{\nu}_{i-1}(z_{[i-1]}) + \frac{1}{\gamma^2}\overline{H}_i^T(z_{[i]})\Delta_i z_i = \frac{1}{\gamma^2}\overline{H}_{[i]}^T(z_{[i]})\Delta_{[i]}z_{[i]}$. To deal with the remaining nonlinearities, the higher order virtual control is selected as

$$\begin{aligned}\overline{\alpha}^h_i &= \overline{G}_i^+\left[A_{\langle i\rangle}(\phi_{[i]}^{-1})^h + f_i^h - \overline{G}^h_i\overline{L}_{\langle i\rangle}z_{[i]} + \overline{L}_{\langle i-1\rangle}\overline{g}^h_{[i-1]}\right. \\ &\quad + \frac{\partial\overline{\alpha}^h_{i-1}}{\partial z_{[i-1]}}\left\{\overline{A}_{[i-1]}z_{[i-1]} + \begin{bmatrix} \overline{g}^h_{[i-2]} \\ \overline{G}_{i-1}z_i \end{bmatrix}\right\} + \frac{\partial\overline{\alpha}_{i-1}}{\partial z_{[i-1]}}\overline{f}^h_{[i-1]} \\ &\quad + \Delta_i^{-1}(\overline{G}^h_{i-1})^T\Delta_{i-1}z_{i-1} + (\overline{H}_i\overline{\nu}_{i-1} - \overline{D}_i\overline{\nu}_{l,i-1}) \\ &\quad \left. + \frac{1}{2\gamma^2}\left(\overline{H}_i\overline{H}_i^T - \overline{D}_i\overline{D}_i^T\right)\Delta_i z_i\right].\end{aligned}$$

Note that all higher order terms in this step, i.e., the terms with superscript "h", follow the definitions that are introduced at the beginning of Section III. Then, we get

$$\dot{\overline{V}}_i = -z_{[i]}^T\overline{Q}_{[i]}z_{[i]} + \gamma^2 w^T w - \gamma^2\|w - \overline{\nu}_i\|^2 + 2z_i^T\Delta_i\overline{G}_i z_{i+1}$$

which is of the form in (14).

Step m: For the last step, we choose $\overline{V}(z) = \overline{V}_{m-1}(z_{[m-1]}) + z_m^T\Delta_m z_m = z^T\Delta z$ as the value function, where z is defined in (12). It is assumed that the dynamics of the $z_{[m-1]}$-subsystem and $\dot{\overline{V}}_{m-1}$ are given by (13) and (14) for $i = m$, respectively. Then, the z_m-dynamics is

$$\begin{aligned}\dot{z}_m &= f_m(x) + G_m(x)u + H_m(x)w \\ &\quad + \frac{\partial\overline{\alpha}_{m-1}}{\partial z_{[m-1]}}\left\{\overline{A}_{[m-1]}z_{[m-1]} + \begin{bmatrix} \overline{g}^h_{[m-2]}(z_{[m-1]}) \\ \overline{G}_{m-1}(z_{[m-1]})z_m \end{bmatrix}\right. \\ &\quad \left. + \overline{H}_{[m-1]}(z_{[m-1]})w + \overline{f}^h_{[m-1]}(z_{[m-1]})\right\} \\ &= \overline{A}_{\langle m\rangle}z + \overline{G}_m(z)u + \overline{H}_m(z)w + \overline{f}^h_m(z),\end{aligned}$$

where

$$\begin{aligned}\overline{f}^h_m(z) &:= A_{\langle m\rangle}(\phi^{-1})^h(z) + f_m^h(z) + \overline{L}_{\langle m-1\rangle}\overline{g}^h_{[m-1]}(z) \\ &\quad + \frac{\partial\overline{\alpha}^h_{m-1}}{\partial z_{[m-1]}}\left\{\overline{A}_{[m-1]}z_{[m-1]} + \begin{bmatrix} \overline{g}^h_{[m-2]}(z_{[m-1]}) \\ \overline{G}_{m-1}(z_{[m-1]})z_m \end{bmatrix}\right\} \\ &\quad + \frac{\partial\overline{\alpha}_{m-1}}{\partial z_{[m-1]}}\overline{f}^h_{[m-1]}(z_{[m-1]}),\end{aligned}$$
$$\overline{H}_m(z) := H_m(\phi^{-1}(z)) + \frac{\partial\overline{\alpha}_{m-1}}{\partial z_{[m-1]}}\overline{H}_{[m-1]}(z_{[m-1]}),$$

and P6 in Lemma 1 is used for $k = m$. As a consequence of the nonlinear *block* backstepping, we have constructed the lower triangular diffeomorphism $z = \phi(x)$ and the CLF $\overline{V}(z) = z^T\Delta z$. Under the coordinate transformation $z = \phi(x)$ and $\overline{G}(z) := [0^T \ \overline{G}_m^T(z)]^T$ with appropriate dimensions, the original *multi-input* nonlinear system (1) is transformed to

$$\dot{z} = \overline{A}z + \overline{G}(z)u + \overline{H}(z)w + \begin{bmatrix} \overline{g}^h_{[m-1]}(z) \\ 0 \end{bmatrix} + \overline{f}^h(z), \tag{15}$$

whose linearized dynamics is the same as (11) since all the quantities are constructed such that the linear parts remain unchanged during the nonlinear backstepping. In particular, the linear part of the diffeomorphism $\phi(x)$ is Lx. This implies that the characteristics of the linearized systems of (1) and (15) are the same which would be a desirable result concerning our dual objectives.

C. Global Inverse Optimal Controller Design

In this subsection, we achieve the main result of the paper, that is, with the CLF $\overline{V}(z) = z^T\Delta z$, a nonlinear controller is designed such that it admits local optimality and global inverse optimality by choosing the positive definite function $\overline{q}(z)$ and designing the positive definite matrix function $\overline{R}(z)$. The detailed descriptions are given in the following theorem.

Theorem 2: There exist a positive definite, radially unbounded function $\overline{q}(z)$ and a matrix $\overline{R}(z)$ with $\overline{q}(0) = 0$, $\overline{q}_z(0) = 0$, $\overline{q}_{zz}(0) = \overline{Q}$, $\overline{R}(z) = \overline{R}^T(z) > 0$ for all z, and $\overline{R}(0) = R_l$ such that for the system (15), the control law

$$u = \overline{\mu}(z) := -\overline{R}^{-1}(z)\overline{G}^T(z)\Delta z \tag{16}$$

achieves local optimality and global inverse optimality with respect to the cost functionals (4) and (8) in z-coordinate, respectively, and the worst case disturbance is

$$w = \overline{\nu}(z) := \frac{1}{\gamma^2}\overline{H}^T(z)\Delta z. \tag{17}$$

Moreover, $z(t) \to 0$ as $t \to \infty$ if $w(t) \in \mathcal{L}_2$, and $z(t) \in \mathcal{L}_\infty$ if $w(t) \in \mathcal{L}_\infty$. If $w \equiv 0$, the origin is globally exponentially stable.

Proof: From the CLF $\overline{V}(z) = z^T\Delta z$ which is obtained at the last step of the nonlinear backstepping procedure, the time derivative of $\overline{V}$ is given by

$$\begin{aligned}\dot{\overline{V}} =& - z^T\overline{Q}z - \gamma^2\overline{\nu}_l^T\overline{\nu}_l + \overline{\mu}_l^T R_l\overline{\mu}_l + \gamma^2\overline{\nu}_{l,m-1}^T\overline{\nu}_{l,m-1} \\ &+ 2\gamma^2\overline{\nu}^T w - \gamma^2\overline{\nu}_{m-1}^T\overline{\nu}_{m-1} + 2z_m^T\Delta_m\overline{G}_m u \\ &+ 2z_m^T\Delta_m\left\{\Delta_m^{-1}(\overline{G}^h_{m-1})^T\Delta_{m-1}z_{m-1} + \overline{f}^h_m\right\},\end{aligned}$$

where $\overline{\nu}$ is defined by (17). Adding and subtracting $u^T\overline{R}(z)u + \overline{\mu}^T\overline{R}(z)\overline{\mu} + \gamma^2 w^T w + \gamma^2\overline{\nu}^T\overline{\nu}$, and completing the squares with respect to u and w yield

$$\begin{aligned}\dot{\overline{V}} =& - z^T\overline{Q}z - u^T\overline{R}(z)u + \gamma^2 w^T w - \gamma^2\|w - \overline{\nu}\|^2 \\ &+ (u - \overline{\mu})^T\overline{R}(z)(u - \overline{\mu}) - \overline{\mu}^T\overline{R}(z)\overline{\mu} + \overline{\mu}_l^T R_l\overline{\mu}_l \\ &+ 2z_m^T\Delta_m\overline{\psi}^h,\end{aligned}$$

where $\overline{\mu}$ is given in (16) and

$$\begin{aligned}\overline{\psi}^h(z) :=& \overline{f}^h_m + \Delta_m^{-1}(\overline{G}^h_{m-1})^T\Delta_{m-1}z_{m-1} + \big(\overline{H}_m\overline{\nu}_{m-1} \\ &-\overline{D}_m\overline{\nu}_{l,m-1}\big) + \frac{1}{2\gamma^2}\left(\overline{H}_m\overline{H}_m^T - \overline{D}_m\overline{D}_m^T\right)\Delta_m z_m,\end{aligned}$$

which consists of only higher order terms. Let

$$\overline{q}(z) := z^T\overline{Q}z + \overline{\mu}^T\overline{R}(z)\overline{\mu} - \overline{\mu}_l^T R_l\overline{\mu}_l - 2z_m^T\Delta_m\overline{\psi}^h.$$

Note that $\overline{q}_{zz}(0) = \overline{Q}$ if $\overline{R}(0) = R_l$. To prove the theorem, we should construct $\overline{R}(z) = \overline{R}^T(z) > 0$ for all z such that $\overline{R}(0) = R_l$ and it renders $\overline{q}(z)$ to be positive definite and radially unbounded.

Since $\overline{\psi}^h(0) = 0$ and $\frac{\partial\overline{\psi}^h}{\partial z}(0) = 0$, there exists $\overline{\Psi}^h(z)$ such that $\overline{\psi}^h(z) = \overline{\Psi}^h(z)z = \overline{\Psi}^h_{[m-1]}(z)z_{[m-1]} + \overline{\Psi}^h_m(z)z_m$ with $\overline{\Psi}^h(0) = 0$. By the Young's inequality, $\overline{q}(z)$ is written by

$$\begin{aligned}\overline{q}(z) =& z^T\overline{Q}z + z_m^T\Delta_m\left\{\overline{G}_m\overline{R}^{-1}(z)\overline{G}_m^T - B_m R_l^{-1}B_m^T\right\} \\ &\times\Delta_m z_m - 2z_m^T\Delta_m\overline{\Psi}^h_{[m-1]}z_{[m-1]} - 2z_m^T\Delta_m\overline{\Psi}^h_m z_m \\ \geq& c\|z\|^2 + z_m^T\Delta_m\Big\{\overline{G}_m\overline{R}^{-1}(z)\overline{G}_m^T - B_m R_l^{-1}B_m^T \\ &-\frac{1}{k}\overline{\Psi}^h_{[m-1]}(\overline{\Psi}^h_{[m-1]})^T - \frac{1}{k}\overline{\Psi}^h_m(\overline{\Psi}^h_m)^T\Big\}\Delta_m z_m,\end{aligned}$$

where $c := \lambda_{\min}(\overline{Q}) - k$ and $\lambda_{\min}(\overline{Q})$ implies the smallest eigenvalue of $\overline{Q}$. If we choose $k < \lambda_{\min}(\overline{Q})$ and take

$$\begin{aligned}\overline{R}^{-1}(z) =& \overline{G}_m^+\Big\{B_m R_l^{-1}B_m^T + \frac{1}{k}\overline{\Psi}^h_{[m-1]}(\overline{\Psi}^h_{[m-1]})^T \\ &+\frac{1}{k}\overline{\Psi}^h_m(\overline{\Psi}^h_m)^T\Big\}(\overline{G}_m^+)^T,\end{aligned} \tag{18}$$

then $\overline{q}(z) \geq c||z||^2 > 0$. Furthermore, $\overline{R}(z) = \overline{R}^T(z) > 0$ for all z, $\overline{R}(0) = R_l$, and $\overline{q}_{zz}(0) = \overline{Q}$ since $\overline{R}^{-1}(z) = \overline{R}^{-T}(z) > 0$ for all z and $\overline{R}^{-1}(0) = R_l^{-1}$. In fact, any matrix $\overline{R}(z) = \overline{R}^T(z) > 0$ for all z such that $\overline{R}(0) = R_l$ and it satisfies (18) with the equality replaced to "$\geq$" can achieve local optimality and global inverse optimality. For the choice of $\overline{R}(z)$, the controller (16) is locally optimal in z-coordinate since $\frac{\partial\overline{\mu}}{\partial z}(0)z = \overline{\mu}_l(z)$. With the designed $\overline{q}(z)$ and $\overline{R}(z)$, the time derivative of the CLF $\overline{V}(z)$ is reduced to

$$\begin{aligned}\dot{\overline{V}} =& - \overline{q}(z) - u^T\overline{R}(z)u + \gamma^2 w^T w - \gamma^2\|w - \overline{\nu}\|^2 \\ &+ (u - \overline{\mu})^T\overline{R}(z)(u - \overline{\mu}).\end{aligned} \tag{19}$$

Thus, the global inverse optimality is guaranteed because (16) and (17) satisfy (9) in z-coordinate, that is,

$$\min_u\max_w\left\{\overline{q}(z) + u^T\overline{R}(z)u - \gamma^2 w^T w + \dot{\overline{V}}(z)\right\} = 0.$$

From this, it is seen that the CLF $\overline{V}(z) = z^T\Delta z$ is actually the value function of the *dynamic game* $\min_u\max_w J(u,w)$ of the *multi-input* nonlinear system (15) for the cost functional (8) in z-coordinate. In addition to this, the optimal control law (16) and the worst case disturbance (17) render (19) to

$$\dot{\overline{V}} = - \overline{q}(z) - u^T\overline{R}(z)u + \gamma^2 w^T w \leq -c||z||^2 + \gamma^2 w^T w,$$

which implies that $z(t) \to 0$ as $t \to \infty$ if $w(t) \in \mathcal{L}_2$ and $z(t) \in \mathcal{L}_\infty$ if $w(t) \in \mathcal{L}_\infty$ (see [7] and [9]). Finally, if $w \equiv 0$, the origin becomes the globally exponentially stable equilibrium point. ■

IV. Conclusion

In this paper, a nonlinear controller for a class of *multi-input* nonlinear systems is proposed. The designed controller guarantees the local optimality and the global inverse optimality. The key tool for the development of the proposed controller is the *block Cholesky factorization*. In the absence of the disturbance, the controller renders the closed-loop system to be globally exponentially stable. This is achieved by hardening the control. If the global exponential stability is not necessary, then the design techniques in [7] can be utilized for less hardened control. Finally, the result of this paper can be extended further to the *multi-input* nonlinear systems that have $\mathcal{C}^1$ (but not smooth) vector fields by using the techniques in [10].

V. Acknowledgments

This work has been supported by Korea Electrical Engineering and Science Research Institute (KESRI, R-2005-7-048), which is funded by the Ministry of Commerce, Industry and Energy (MOCIE), Korea.

References

[1] H.K. Khalil, *Nonlinear Systems*, 3rd Ed., Prentice-Hall, 2002.
[2] R. Sepulchre, M. Jankovic, and P.V. Kokotović, *Constructive Nonlinear Control*, Springer-Verlag, 1997.
[3] M. Krstić, I. Kanellakopoulos, and P.V. Kokotović, *Nonlinear and Adaptive Control Design*, Wiley, New York, 1995.
[4] T. Basar and P. Bernhard, *$\mathcal{H}_\infty$-Optimal Control and Related Minimax Design Problems*, Birkhaüser, Boston, 1991.
[5] K. Zhou, J.C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, New Jersey, 1996.
[6] E. Sontag, "A 'Universal' Construction of Arstein's Theorem on Nonlinear Stabilization," *Systems and Control Letters*, vol. 13, pp. 117–123, 1989.
[7] K. Ezal, Z. Pan, and P.V. Kokotović, "Locally Optimal and Robust Backstepping Design," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 260–271, 2000.
[8] J.P. Strand, K. Ezal, T.I. Fossen, and P.V. Kokotović, "Nonlinear Control of Ships: A Locally Optimal Design," *in Proc. Nonlinear Contr. Syst. Design Symp.*, pp. 732–737, 1998.
[9] A.R. Teel, "Asymptotic Convergence from $\mathcal{L}_p$ Stability," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 2169–2170, 1999.
[10] S. Kang, J. Back, H. Shim, and J.H. Seo, "Locally Optimal and Robust Backstepping Design for $\mathcal{C}^1$ Vector Fields," *in Proc. International Conference on Control, Automation and Systems*, 2007.

Appendix
Block Cholesky Factorization

The conventional *block Cholesky factorization* is of the form $P = LL^T$, where L is block lower triangular matrix. On the other hand, it is necessary in this paper to factorize $P = L^T\Delta L$, where L and Δ are block lower triangular and diagonal matrices, respectively. Thus, we give the modified version of the conventional *block Cholesky factorization* for the completeness of the paper. The following two properties of a block matrix will be used for the proof of the forthcoming lemma.

M1: If L is a block lower/upper triangular matrix with identity matrices along the diagonal, then its inverse L^{-1} is of the same form, that is, L^{-1} is also a block lower/upper triangular matrix with identity matrices, which are of the same dimensions as the ones in L, along the diagonal.

M2: For any given symmetric positive definite matrix P, $P_{[i]}$ is also symmetric positive definite for all i.

Lemma 3: For any given symmetric positive definite matrix $P \in \mathbb{R}^{n\times n}$ and positive integers μ_i for $i = 1, \ldots, m$ where $\mu_1+\cdots+\mu_m = n$, there exists the unique factorization $P = L^T\Delta L$ such that

$$L = \begin{bmatrix} I_1 & 0 & \cdots & 0 \\ L_{11} & I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{m-1,1} & L_{m-1,2} & \cdots & I_m \end{bmatrix},$$
$$\Delta = \mathrm{diag}(\Delta_1, \Delta_2, \cdots, \Delta_m),$$

where $\Delta_i,\ I_i \in \mathbb{R}^{\mu_i\times\mu_i}$, $\Delta_i = \Delta_i^T > 0$ for all i, and other matrices are of appropriate dimensions.

Proof: Define $S := P^{-1}$. We will seek the factorization $S = U^T DU$, where U is a block upper triangular matrix and D is a block diagonal matrix.

For $i = 1$, define $U_{[1]} := I_1$ and $D_{[1]} := S_{[1]} = S_{11}$. Then $S_{[1]} = U_{[1]}^T D_{[1]} U_{[1]}$ with $D_{[1]} = D_{[1]}^T > 0$ by M2.

For $i > 1$, suppose that $S_{[i-1]} = U_{[i-1]}^T D_{[i-1]} U_{[i-1]}$, where $U_{[i-1]}$ and $D_{[i-1]}$ have the same form as in $L_{[i-1]}^T$ and $\Delta_{[i-1]}$, respectively, and $D_j = D_j^T > 0$ for $j = 1, \ldots, i-1$. We postmultiply a block upper triangular matrix on $S_{[i]}$ to obtain the expression like below:

$$\begin{bmatrix} S_{[i-1]} & \widehat{S}_i \\ \widehat{S}_i^T & S_{ii} \end{bmatrix} \begin{bmatrix} U_{[i-1]}^{-1} & -S_{[i-1]}^{-1}\widehat{S}_i \\ 0 & I_i \end{bmatrix}$$
$$= \begin{bmatrix} U_{[i-1]}^T & 0 \\ \widehat{S}_i^T U_{[i-1]}^{-1} D_{[i-1]}^{-1} & I_i \end{bmatrix} \begin{bmatrix} D_{[i-1]} & 0 \\ 0 & S_{ii} - \widehat{S}_i^T S_{[i-1]}^{-1}\widehat{S}_i \end{bmatrix}$$

where $\widehat{S}_i := [S_{1i}^T\ \cdots\ S_{i-1,i}^T]^T$. Postmultiply the inverse of the block upper triangular matrix on the above equation, then

$$S_{[i]} = \begin{bmatrix} U_{[i-1]}^T & 0 \\ \widehat{S}_i^T U_{[i-1]}^{-1} D_{[i-1]}^{-1} & I_i \end{bmatrix} \begin{bmatrix} D_{[i-1]} & 0 \\ 0 & S_{ii} - \widehat{S}_i^T S_{[i-1]}^{-1}\widehat{S}_i \end{bmatrix}$$
$$\times \begin{bmatrix} U_{[i-1]} & D_{[i-1]}^{-1} U_{[i-1]}^{-T}\widehat{S}_i \\ 0 & I_i \end{bmatrix}$$
$$=: U_{[i]}^T D_{[i]} U_{[i]}.$$

Let $D_i := S_{ii} - \widehat{S}_i^T S_{[i-1]}^{-1}\widehat{S}_i$, then $D_i = D_i^T > 0$ since $S_{[i]}$ is positive definite by M2 and D_i is the Schur complement of the matrix $S_{[i]}$. Thus, S can be factored as $S = U^T DU$ which is the case for $i = m$. By letting $L := U^{-T}$ and $\Delta := D^{-1}$, P can be written as $P = L^T\Delta L$, where L and Δ are the ones in Lemma 3. To show the uniqueness of the factorization, suppose that P has two factorizations, namely, $P = L_1^T\Delta_1 L_1 = L_2^T\Delta_2 L_2$. By premultiplying L_1^{-T} and postmultiplying L_2^{-1} on the equation, we have $\Delta_1 L_1 L_2^{-1} = L_1^{-T} L_2^T \Delta_2$. The left hand side of the equation is a block lower triangular matrix with the diagonals are in Δ_1 and the right hand side is a block upper triangular matrix with the diagonals are in Δ_2. Therefore, $\Delta_1 = \Delta_2$ and $L_1 L_2^{-1} = I$. ∎