Networked and Distributed Control Systems

(SC42100)

Assignment 3

Responsible Instructors

Student

Daniel Varela - 5321263

Tamás Keviczky

Manuel Mazo

Gabriel Gleizer

June 29, 2021

Contents

Introduction

This Assignment looks at some theoretical results useful in the implementation of the algorithms and methods discussed in class for distributed optimization techniques, giving further insight into these methods.

Section 1 looks at a simple primal decomposition example and follows up with an important result on the relation between subgradients and dual optimal variables. In Section 2, the attention turns to the combined consensus/projected incremental subgradient method, where equivalence is shown between this method and the standard subgradient method when a large enough number of consensus iterations is performed. Finally, Sections 3 and 4 look at two problems based of the Model Predictive Control: Theory, Computation, and Design book [Rawlings and Mayne, 2008] to derive some results pertaining to the optimization of a set of variables one at a time.

1 Problem 1

This section covers concepts related to primal and dual decomposition.

1.1 Primal Decomposition

The starting point is the convex optimization problem with a complicating constraint:

$$\begin{array}{ll} \underset{\theta_1,\theta_2}{\text{minimize}} & f_1(\theta_1) + f_2(\theta_2) \\ \text{subject to} & \theta_1 \in \Theta_1, \quad \theta_2 \in \Theta_2 \\ & h_1(\theta_1) + h_2(\theta_2) \leq 0 \end{array} \tag{1}$$

where Θ_1, Θ_2 are convex sets and all the functions are convex. Primal Decomposition is usually done for problem of the following form:

$$\begin{array}{ll} \underset{y,\{\theta^i\}}{\text{minimize}} & \sum_i f^i\left(\theta^i\right) \\ \text{subject to} & \theta^i \in \Theta^i, \forall i \\ & A^i\theta^i \leq y, \forall i \\ & y \in \mathcal{Y} \end{array} \tag{2}$$

As such, applying primal decomposition to 1 can be done by rewritting the constraints to fit structure of problem 2. This is done easily by introducing a variable $r \in \mathbb{R}$ such that:

$$\begin{aligned} \underset{\theta_1,\theta_2}{\text{minimize}} \quad & f_1(\theta_1) + f_2(\theta_2) \\ \text{subject to} \quad & \theta_1 \in \Theta_1, \quad \theta_2 \in \Theta_2 \\ & h_1(\theta_1) \leq r \\ & h_2(\theta_2) \leq -r \end{aligned} \tag{3}$$

Note that $h_1(\theta_1) + h_2(\theta_1) \leq r - r = 0$, so the initial condition still holds.

The problem can now be split into two lower order problems:

$$\begin{aligned} \underset{\theta_1}{\text{minimize}} \quad & f_1(\theta_1) & \qquad \underset{\theta_2}{\text{minimize}} \quad & f_2(\theta_2) \\ \text{subject to} \quad & \theta_1 \in \Theta_1 & \qquad \text{subject to} \quad & \theta_2 \in \Theta_2 \\ & h_1(\theta_1) \leq r & & h_1(\theta_2) \leq -r \end{aligned} \tag{4}$$

Which effectively separate the optimization variables θ_1 and θ_2, which are solved for independently. Compatibilization of both solutions can be done by introducing a higher level problem:

$$\begin{aligned} \underset{r}{\text{minimize}} \quad & \phi_1(r) + \phi_2(r) \\ \text{with} \quad & \phi_1(r) = \inf_{\theta_1} f_1(\theta_1) \ : \ \theta_1 \leq r \\ & \phi_2(r) = \inf_{\theta_2} f_2(\theta_2) \ : \ \theta_2 \leq -r \end{aligned} \tag{5}$$

1.2 Subgradient and Optimal Dual Variables

Solving the master problem above requires that each individual sub-problem returns its respective subgradient. This can be found from an optimal dual variable associated with the coupling constraint. To prove this finding, the basis is a paper by Björn Johansson, [Johansson et al., 2006]. Start by considering the problem, with optimal value $p(z)$.

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & f(\theta) \\ \text{subject to} \quad & \theta \in \Theta \\ & h(\theta) \leq z \end{aligned} \tag{6}$$

The aim is to prove that, for an optimal dual variable λ^* associated to the constraint $h(\theta) \leq z$, $-\lambda^*$ is a subgradient of p at z.

The dual problem associated to the formulation in 6 is:

$$q(z) = \max_{\lambda \geq 0} d(\lambda, z) = \max_{\lambda \geq 0} \min_{\theta \in \Theta} L(\theta, z, \lambda) \ , \quad L(\theta, z, \lambda) = f(\theta) + \lambda^{\intercal}(h(\theta) - z) \tag{7}$$

Under the assumption that Θ is a convex set, $h(\theta)$ is convex, and that there exists a feasible $\bar{\theta}$ such that $h(\bar{\theta}) < z$, then strong convexity will hold by Slater's theorem. Then, the duality gap is null and the optimal cost of the dual problem will be the same as the optimal cost of the primal problem.

$$p(z) = q(z) = \max_{\lambda \geq 0} d(\lambda, z) \tag{8}$$

Take two feasible points, z and y, and λ^*, the Lagrange Multipliers corresponding to the constraint $h(\theta) - z \leq 0$. Since λ^* is optimal when associated to z and not necessarily y, then:

$$\begin{aligned}
p(y) =& \max_{\lambda_i} \left\{ \min_{\theta} \{ f(\theta) + \lambda^\intercal (h(\theta) - y) \} \right\} \\
\geq& \min_{\theta} \{ f(\theta) + (\lambda^*)^\intercal (h(\theta) - y) \} \\
=& \min_{\theta} \{ f(\theta) + (\lambda^*)^\intercal (h(\theta) - z) \} - (\lambda^*)^\intercal (y - z) \\
=& p(z) - (\lambda^*)^\intercal (y - z)
\end{aligned} \tag{9}$$

Hence, by definition of a subgradient, $-\lambda^*$ is a subgradient of $p(\cdot)$ at z.

2 Problem 2 - Combined Consensus/Incremental Subgradient

In this section, a proof will be given on how the the combined consensus / projected incremental subgradient method becomes a standard subgradient method as the number of consensus iterations increases to infinity. Before getting to the actual subgradient methods, a brief walk through the properties of doubly stochastic matrices is given and all the mathematical background is laid out, to make the comparison easier.

Note that this section is heavily inspired by two other papers. [Olfati-Saber et al., 2007] is a main inspiration for Subsection 2.1. [Johansson et al., 2008] is heavily cited in Subsection 2.2.

2.1 Properties of Doubly Stochastic Matrices

2.1.1 Definition

A doubly stochastic matrix, is a square matrix $W = (w_{ij})$ of size N, $W \in \mathbb{R}^{N \times N}$, of non-negative real numbers, whose rows and columns all sum up to 1.

$$\sum_{i=1}^{N} w_{ij} = \sum_{j=1}^{N} w_{ij} = 1 \tag{10}$$

2.1.2 Eigenvalues

The eigenvalues of doubly stochastic matrices have some properties that can be of use in this scenario.

For starters, $\lambda_1 = 1$ is an eigenvalue of W. Note simply that all eigenvalues of W must meet the property $Wv_i = \lambda_i v_i$, where v_i is the right eigenvector associated to eigenvalue λ_i. Suppose $e_i = \mathbf{1}_{N\times 1}$, a column vector of size $N \times 1$ with all entries equal to 1. The result of the multiplication of W with this vector is simply the sum of each of the rows in W. Hence, $W\mathbf{1}_{N\times 1} = \lambda_i \mathbf{1}_{N\times 1}$ holds trivially for the eigenvalue $\lambda_i = 1$.

From application of the Gershgorin circle theorem, W will have no eigenvalues outside the unit circle, so it follows that the largest eigenvalue of W is $\lambda_1 = 1$.

Since the graph being used is strongly connected, this means that the underlying matrix W is irreducible. Before continuing, one assumption is needed here. As such, we proceed as in [Johansson et al., 2008] and [Olfati-Saber et al., 2007] and assume the eigenvalue $\lambda_1 = 1$ is the unique largest real eigenvalue of W, meaning $1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_N$.

2.1.3 Decomposition

Matrix W admits the following decomposition:

$$W = SJS^{-1} \tag{11}$$

where S is a matrix whose columns are the right eigenvectors of W, S^{-1} a matrix whose rows are the left eigenvectors of W and J is a Jordan matrix with the (possibly repeated) eigenvalues of W along its main diagonal. For a total of N eigenvalues, with n distinct eigenvalues, the matrix will look generically like:

$$J = \begin{bmatrix} [\,1\,] & & & \\ & \begin{bmatrix} \lambda_2 & 1 & & \\ & & \ddots & 1 \\ & & & \lambda_2 \end{bmatrix} & & \\ & & \ddots & \\ & & & \begin{bmatrix} \lambda_n & 1 & & \\ & & \ddots & 1 \\ & & & \lambda_n \end{bmatrix} \end{bmatrix} \tag{12}$$

Let $J = \text{diag}\{1, J_2, \dots J_n\}$, where J_i represents the Jordan block associated to the $i-th$ distinct eigenvalue of W. Then similarly, matrices S and S^{-1} accept the following partition

$$\begin{cases} S = [v_1, v_{J_2}, \dots v_{J_n}], \text{ the right eigenvectors of } W \text{ associated to block } J_i \\ S^{-1} = [u_1^\intercal, u_{J_2}^\intercal, \dots u_{J_n}^\intercal]^\intercal, \text{ the left eigenvectors of } W \text{ associated to block } J_i \end{cases} \tag{13}$$

Then matrix W can be written as:

$$W = v_1 \lambda_1 u_1 + \sum_{i=2}^{n} v_{J_i} J_i u_{J_i} = v_1 u_1 + \sum_{i=2}^{n} v_{J_i} J_i u_{J_i} \tag{14}$$

2.1.4 Powers

Taking powers of W can be done with the help of the decomposition above.

$$W^\varphi = S J^\varphi S^{-1} \tag{15}$$

where $J^\varphi = \text{diag}\{1^\varphi, J_2^\varphi, \dots J_n^\varphi\}$. The powers of a Jordan block J_i are given generically by:

$$J_i^\varphi = \begin{bmatrix} \binom{\varphi}{0}\lambda_i^\varphi & \binom{\varphi}{1}\lambda_i^{\varphi-1} & \binom{\varphi}{2}\lambda_i^{\varphi-2} & \cdots & \binom{\varphi}{N-1}\lambda_i^{\varphi-N+1} \\ 0 & \binom{\varphi}{0}\lambda_i^\varphi & \binom{\varphi}{1}\lambda_i^{\varphi-1} & \cdots & \binom{\varphi-2}{N-2}\lambda_i^{\varphi-N+2} \\ 0 & 0 & \binom{\varphi}{0}\lambda_i^\varphi & \cdots & \binom{\varphi-3}{N-3}\lambda_i^{\varphi-N+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{\varphi}{0}\lambda_i^\varphi \end{bmatrix} \tag{16}$$

Since all eigenvalues of W which are different from $\lambda_1 = 1$ have that $\|\lambda_i\| < 1$, then it follows that:

$$\lim_{\varphi \to \infty} J_i^\varphi = 0 \tag{17}$$

As such, one can use the decomposition in equation 14 to show that $W^\varphi = v_1 u_1$ if $\varphi \to \infty$.

$$\begin{aligned} W^\varphi &= v_1 \lambda_1^\varphi u_1 + \sum_{i=2}^{n} v_{J_i} J_i^\varphi u_{J_i} \\ \lim_{\varphi \to \infty} W^\varphi &= \lim_{\varphi \to \infty} v_1 1^\varphi u_1 + \lim_{\varphi \to \infty} \sum_{i=2}^{n} v_{J_i} J_i u_{J_i} \\ \lim_{\varphi \to \infty} W^\varphi &= v_1 u_1 \end{aligned} \tag{18}$$

Further, since $v_1 = \mathbf{1}_{N\times 1}$, matrix W^∞ will just be a matrix whose rows are all the same and equal to u_1:

$$W^\infty = \begin{bmatrix} u_1^\intercal & u_1^\intercal & \dots & u_1^\intercal \end{bmatrix}^\intercal \tag{19}$$

2.1.5 Choice of Left and Right Eigenvectors

The choice of vector u_1 is important here. For instance, in the same way that $W\mathbf{1}_{N\times1} = \mathbf{1}_{N\times1}$, it also holds that $\mathbf{1}_{1\times N}W = \mathbf{1}_{1\times N}$. In general, all vectors of the type $\alpha\mathbf{1}_{N\times1}$ with $\alpha \in \mathbb{R}$, can be both seen as left and right eigenvectors of W corresponding to the eigenvalue $\lambda_1 = 1$. To make a choice on the correct length of the eigenvalues, let $W^2 = M = (m_{ij})$. Given that W is doubly stochastic, equation 10 holds. Then, it follows that:

$$\begin{aligned}
\sum_j m_{ij} &= \sum_j \left(\sum_k w_{ik} w_{kj} \right) = \sum_j \sum_k w_{ik} w_{kj} = \sum_k \sum_j w_{ik} w_{kj} \\
&= \sum_k w_{ik} \left(\sum_j w_{kj} \right) = \sum_k w_{ik}(1) = 1 \\
\sum_i m_{ij} &= \sum_i \left(\sum_k w_{kj} w_{ik} \right) = \sum_i \sum_k w_{kj} w_{ik} = \sum_k \sum_i w_{jk} w_{ik} \\
&= \sum_k w_{kj} \left(\sum_i w_{ik} \right) = \sum_k w_{kj}(1) = 1
\end{aligned} \tag{20}$$

The implication is that the $M = W^2$ is also a doubly stochastic matrix. In fact, in general it holds that the product of doubly stochastic matrices is also doubly stochastic, and the proof shown above holds both for generic matrices and for higher powers of W.

The important implication here is that W^∞ should also be doubly stochastic. This further implies that v_1 and u_1 should be chosen such that $u_1 v_1 = 1$. If the choice is made to keep $v_1 = \mathbf{1}_{N\times1}$, then there's an unique value of α that preserves the stochasticity properties of W^∞, and the conclusion is that $u_1 = \frac{1}{N}\mathbf{1}_{1\times N}$. Finally, one can arrive at a value for W^∞:

$$\lim_{\varphi\to\infty} W^\varphi = v_1 u_1 = \frac{1}{N}\mathbf{1}_{N\times1}\mathbf{1}_{1\times N} \tag{21}$$

2.2 Subgradient Methods

The combined consensus/projected incremental subgradient method for N agents is given in the lecture slides as:

$$\theta^i_{k+1} = \mathcal{P}_\Theta \left[\sum_{j=1}^{N} [W^\varphi]_{ij} \left(\theta^j_k - \alpha_k g^j \left(\theta^j_k \right) \right) \right], \quad i = 1, \ldots, N \tag{22}$$

and aims at the use of agreement protocols to relax communication constraints in distributed optimization schemes. Note that each agent updates it's own internal copy of θ_k, by performing a local subgradient update. All of the individual updates are then combined with recourse to the consensus matrix W, whose goal is to compatibilize the different estimates of different agents according to the underlying communication graph.

By contrast, the standard subgradient method merely updates by summing all the subgradient updates of all N agents:

$$\theta_{k+1} = \mathcal{P}_\Theta \left[\theta_k - \beta_k \sum_{i=1}^{N} g^i(\theta_k) \right] \tag{23}$$

Returning to the projected incremental subgradient method, note that, as shown in the previous subsection, taking $\varphi \to \infty$, the method becomes:

$$\theta_{k+1}^i = \mathcal{P}_\Theta \left[\sum_{j=1}^{N} \frac{1}{N} \left(\theta_k^j - \alpha_k g^j \left(\theta_k^j \right) \right) \right] = \mathcal{P}_\Theta \left[\frac{1}{N} \sum_{j=1}^{N} \left(\theta_k^j - \alpha_k g^j \left(\theta_k^j \right) \right) \right] \tag{24}$$

Note that the iteration of all agents will be the same, simply corresponding to the average of all the agent's local values. This further implies that, in all iterations going forward, all the local copies will be the same. As such, one can rewrite this projection method as:

$$\theta_{k+1}^i = \mathcal{P}_\Theta \left[\frac{1}{N} \sum_{j=1}^{N} \left(\theta_k^j - \alpha_k g^j \left(\theta_k^j \right) \right) \right] \Leftrightarrow \theta_{k+1} = \mathcal{P}_\Theta \left[\theta_k - \frac{\alpha_k}{N} \sum_{j=1}^{N} g^j \left(\theta_k \right) \right] \tag{25}$$

This is nothing but the standard subgradient method, with step-size $\beta_k = \frac{\alpha_k}{N}$. The conclusion is that, as $\varphi \to \infty$, the combined consensus / projected incremental subgradient method becomes a standard subgradient method.

3 Problem 3 - Optimizing One Variable at a Time

This section is concerned with the optimization of a quadratic value function by dividing the set of variables in two and optimizing each set of variables based on the optimal value obtained for the other set, iteratively. The value function is defined as:

$$\begin{aligned} V(u) &= \frac{1}{2} u^\intercal H u + c^\intercal u + d \\ V(u_1, u_2) &= \frac{1}{2} \begin{pmatrix} u_1^\intercal & u_2^\intercal \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} c_1^\intercal & c_2^\intercal \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + d \end{aligned} \tag{26}$$

3.1 Problem 3.1

In showing that the next points in the iteration are:

$$\begin{aligned} u_1^{p+1} &= -H_{11}^{-1} \left(H_{12} u_2^p + c_1 \right) \\ u_2^{p+1} &= -H_{22}^{-1} \left(H_{21} u_1^p + c_2 \right) \end{aligned} \tag{27}$$

One can start by understanding that the points u_1^{p+1} and u_2^{p+1} are merely the optimal values u_1^* and u_2^* resulting from the optimization (minimization, specifically) of the value function in

26 with respect to u_1^p and u_2^p, respectively. Note additionally that, since the value function is quadratic in nature, these optimal points can easily be obtained by taking the partial derivatives of $V(u)$ with respect to each variable, and setting them to zero, thus finding the point at which the function is minimized.

$$\begin{aligned} V(u_1, u_2) =& \frac{1}{2} \begin{pmatrix} u_1^\intercal & u_2^\intercal \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} c_1^\intercal & c_2^\intercal \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + d \\ =& \frac{1}{2} \left(u_1^\intercal H_{11} u_1 + u_2^\intercal H_{21} u_1 + u_1^\intercal H_{12} u_2 + u_2^\intercal H_{22} u_2 \right) + c_1^\intercal u_1 + c_2^\intercal u_2 + d \end{aligned} \tag{28}$$

Taking the partial derivatives with respect to each set of variables:

$$\begin{aligned} & \frac{\partial}{\partial u_1} \left(V(u_1, u_2) \right) \\ =& \frac{1}{2} \left(u_1^\intercal H_{11} + H_{11} u_1 + u_2^\intercal H_{21} + H_{12} u_2 \right) + c_1 = \frac{1}{2} \left(2H_{11} u_1 + 2H_{12} u_2 \right) + c_1 \\ =& H_{11} u_1 + H_{12} u_2 + c_1 \end{aligned} \tag{29}$$

$$\begin{aligned} & \frac{\partial}{\partial u_2} \left(V(u_1, u_2) \right) \\ =& \frac{1}{2} \left(u_2^\intercal H_{22} + H_{22} u_2 + u_1^\intercal H_{12} + H_{21} u_1 \right) + c_2 = \frac{1}{2} \left(2H_{22} u_2 + 2H_{21} u_1 \right) + c_2 \\ =& H_{22} u_2 + H_{21} u_1 + c_2 \end{aligned} \tag{30}$$

Equating these to zero:

$$\begin{aligned} \frac{\partial}{\partial u_1} \left(V(u_1, u_2) \right) = 0 \implies u_1^* = -H_{11}^{-1} \left(H_{12} u_2 + c_1 \right) \\ \frac{\partial}{\partial u_2} \left(V(u_1, u_2) \right) = 0 \implies u_2^* = -H_{22}^{-1} \left(H_{21} u_1 + c_2 \right) \end{aligned} \tag{31}$$

These two equations can easily be written in matrix form as:

$$u^{p+1} = \begin{bmatrix} 0 & -H_{11}^{-1} H_{12} \\ -H_{22}^{-1} H_{21} & 0 \end{bmatrix} u^p + \begin{bmatrix} -H_{11}^{-1} c1 \\ -H_{22}^{-1} c2 \end{bmatrix} \Leftrightarrow u^{p+1} = Au^p + b \tag{32}$$

3.2 Problem 3.2

The goal of this section is to provide formal justification that the following holds:

$$|\operatorname{eig}(A)| < 1 \tag{33}$$

The solution to this problem is direct by application of the Householder-John theorem. Since the solution is quite simple, it is briefly presented in the following.

The starting point is the definition of eigenvalues for finite dimensional matrices. Let λ be an eigenvalue of A and v the corresponding right eigenvector. Then it holds that:

$$Av = \lambda v \tag{34}$$

The first step is realize that:

$$A = \begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & -H_{11} & 0 \\ 0 & -H_{22} & \end{bmatrix}^{-1} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = I - D^{-1}H \quad (35)$$

Developing equation 34:

$$\begin{gathered} Av = \lambda v \Leftrightarrow (I - D^{-1}H)v = \lambda v \Leftrightarrow v - D^{-1}Hv = \lambda v \\ (1-\lambda)v = D^{-1}Hv \Leftrightarrow (1-\lambda)Dv = Hv \end{gathered} \quad (36)$$

Left multiplying both sides with $v^\intercal$:

$$v^\intercal(1-\lambda)Dv = v^\intercal Hv \Leftrightarrow v^\intercal Dv = \frac{1}{1-\lambda}v^\intercal Hv \quad (37)$$

One can already realize something about λ. Note that both D and H are positive definite matrices, and therefore both quantities $v^\intercal Dv > 0$ and $v^\intercal Hv > 0$. This naturally implies that $(1-\lambda)^{-1} > 0 \Leftrightarrow \lambda < 1$. The proof, however, is not yet complete. Consider given hint:

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \succ 0 \Rightarrow \bar{H} = \begin{pmatrix} H_{11} & -H_{12} \\ -H_{21} & H_{22} \end{pmatrix} \succ 0 \quad (38)$$

Note additionally that $\bar{H} = 2D - H$. Then, it's possible to use of the equality in 37 to write:

$$v^\intercal(2D - H)v = \left(\frac{1}{1-\lambda} + \frac{1}{1-\lambda} - 1\right)v^\intercal Hv \Leftrightarrow v^\intercal \bar{H}v = \left(\frac{1+\lambda}{1-\lambda}\right)v^\intercal Hv \quad (39)$$

Applying the same reasoning as before:

$$\left(\frac{1+\lambda}{1-\lambda}\right) > 0 \Leftrightarrow \begin{cases} (1+\lambda > 0) \wedge (1-\lambda > 0) \\ (1+\lambda < 0) \wedge (1-\lambda < 0) \end{cases} \implies \lambda > 1 \wedge \lambda < 1 \implies |\lambda| < 1 \quad (40)$$

The implication here is that any eigenvalue of the original matrix A needs to fulfill the above condition. The proof is complete and the proposition holds:

$$|\operatorname{eig}(A)| < 1 \quad (41)$$

3.3 Problem 3.3

Showing convergence of the process to its optimal value u^* is the same as showing that, at the limit, u^{p+1} is no longer different from u^p, implying the optimal point has been reached. Substituting in $u^{p+1} = u^p = u*$ in the dynamical equation found in 32:

$$\begin{gathered} u^* = Au^* + b \Leftrightarrow (I - A)u^* = b \Leftrightarrow (I-A)u^* = \begin{bmatrix} -H_{11}^{-1} & 0 \\ 0 & -H_{22}^{-1} \end{bmatrix} \begin{bmatrix} c1 \\ c2 \end{bmatrix} \\ \Leftrightarrow \begin{bmatrix} -H_{11} & 0 \\ 0 & -H_{22} \end{bmatrix} \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix} u^* = c \Leftrightarrow -Hu^* = c \implies u^* = -H^{-1}c \end{gathered} \quad (42)$$

4 Problem 4 - Monotonically Decreasing Cost

4.1 Problem 4.1

To prove the monotonic decrease of the value function, the basis will be a result obtained in Section 4.2. Summarily, the result shows that:

$$V(u^{p+1}) - V(u^p) = -\frac{1}{2}(u^p - u^*)^\intercal P(u^p - u^*), \quad P = HD^{-1}\tilde{H}D^{-1}H, \quad \tilde{H} = D - N \tag{43}$$

Hence, to prove the monotonic decrease of the value function, one need only prove that the matrix P shown in the expression is positive definite for all $u^p \neq -H^{-1}c$. This would make the whole expression negative definite, on account of the minus sign, thus proving that the value function is monotonically decreasing.

$$V(u^{p+1}) < V(u^p) \forall u^p \neq -H^{-1}c \tag{44}$$

One can initially show that P can be decomposed into a specific structure, $P = R^\intercal QR$, with $R = D^{-1}H$ and $Q = \tilde{H} \succ 0$. Note initially that:

$$(D^{-1}H)^\intercal = \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix}^\intercal = \begin{bmatrix} I & (H_{22}^{-1}H_{21})^\intercal \\ (H_{11}^{-1}H_{12})^\intercal & I \end{bmatrix} = \begin{bmatrix} I & H_{12}H_{22}^{-1} \\ H_{12}H_{11}^{-1} & I \end{bmatrix} = HD^{-1} \tag{45}$$

This result holds since, for block partitioned matrices, the following is true:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad M^\intercal = \begin{bmatrix} A^\intercal & C^\intercal \\ B^\intercal & D^\intercal \end{bmatrix} \tag{46}$$

Further, since the original matrix H is symmetric and positive definite, then H_{11}^{-1} and H_{22}^{-1} are also symmetric and positive definite, $H_{11}^{-1} = H_{11}^{-1\intercal}$ and $H_{22}^{-1} = H_{22}^{-1\intercal}$. Further, symmetry of H also implies $H_{12}^\intercal = H_{21}$ and $H_{21}^\intercal = H_{12}$.

Proving the non-singularity of $R = D^{-1}H$ is achievable by considering that R is the product of two positive definite matrices.

$$R = I - H_{11}^{-1}H_{12}H_{22}^{-1}H_{21} = H_{11}^{-1}\left(H_{11} - H_{12}H_{22}^{-1}H_{21}\right) \tag{47}$$

Since the determinant of positive definite matrices is positive, and a property of determinants is that $\det(AB) = \det(A)\det(B)$:

$$\begin{cases} \det(H_{11}^{-1}) > 0 \\ \det\left(H_{11} - H_{12}H_{22}^{-1}H_{21}\right) > 0 \end{cases} \implies \det(I - H_{11}^{-1}H_{12}H_{22}^{-1}H_{21}) = \det(R) > 0 \tag{48}$$

It follows that, since $\det(R) > 0$, R is also invertible, thus proving its non-singularity.

Making use of the hint in equation 38, it's easy to see that $Q = \tilde{H} \succ 0$. The final step is to apply the second hint:

$$Q > 0 \text{ and } R \text{ non-singular } \Rightarrow R^T Q R > 0 \tag{49}$$

To show that:

$$P = HD^{-1}\tilde{H}D^{-1}H \succ 0 \tag{50}$$

Given that $P \succ 0$, the assumption about the negative definiteness of $V(u^{p+1}) - V(u^p)$ for all $u^p \neq u^*$ is direct and thus:

$$V\left(u^{p+1}\right) < V\left(u^p\right) \quad \forall u^p \neq -H^{-1}c \tag{51}$$

4.2 Problem 4.2

The approach followed in this section was to find to develop both expressions and equate them to each other, in order to find equivalences between the terms on each side.

To that end, one can start by developing the left hand side of the equation.

$$V(u^p) = \frac{1}{2}u^{p\intercal}Hu^p + c^\intercal u^p + dV(u^{p+1}) = \frac{1}{2}u^{p+1\intercal}Hu^{p+1} + c^\intercal u^{p+1} + d \tag{52}$$

With information about the dynamics of the system, one can replace u^{p+1} in the second equation:

$$V(u^{p+1}) = \frac{1}{2}(Au^p + b)^\intercal H\left(Au^p + b\right) + c^\intercal\left(Au^p + b\right) + d \tag{53}$$

The difference between the two terms can now be expressed in terms of u^p.

$$\begin{aligned} &V(u^{p+1}) - V(u^p) \\ =&\frac{1}{2}\left(u^p\left(A^\intercal HA - H\right)u^{p\intercal} + b^\intercal HAu^p + u^{p\intercal}A^\intercal Hb + b^\intercal Hb\right) + c^\intercal\left(A - I\right)u^p + c^\intercal b \end{aligned} \tag{54}$$

Note further that, since the linear term $c^\intercal(A - I)u^p$ is a scalar, one can decompose it as:

$$c^\intercal(A - I)u^p = \frac{1}{2}c^\intercal(A - I)u^p + \frac{1}{2}u^{p\intercal}(A^\intercal - I)c \tag{55}$$

Resulting in:

$$\begin{aligned} &V(u^{p+1}) - V(u^p) \\ =&\underbrace{\frac{1}{2}u^p\left(A^\intercal HA - H\right)u^{p\intercal}}_{\Delta V_1} + \underbrace{\frac{1}{2}\left(b^\intercal HAu^p + c^\intercal\left(A - I\right)\right)u^p}_{\Delta V_2} + \underbrace{\frac{1}{2}u^{p\intercal}\left(A^\intercal Hb + (A^\intercal - I)c\right)}_{\Delta V_3} + \underbrace{\frac{1}{2}b^\intercal Hb + c^\intercal b}_{\Delta V_4} \end{aligned} \tag{56}$$

Equating this to the right-hand side of the equation, it's easy to find the match between terms quadratic, linear and independent of u^p.

$$\begin{aligned}&-\frac{1}{2}(u^p-u^*)^\intercal P(u^p-u^*)=\\&=\underbrace{-\frac{1}{2}u^{p\intercal}Pu^p}_{\Delta V_1}+\underbrace{\frac{1}{2}u^{p\intercal}Pu^*}_{\Delta V_2}+\underbrace{\frac{1}{2}u^{*\intercal}Pu^p}_{\Delta V_3}\underbrace{-\frac{1}{2}u^{*\intercal}Pu^*}_{\Delta V_4}\end{aligned}\tag{57}$$

Before continuing, note that the original dynamical equation can be easily rewritten in terms of matrices D, N and c:

$$A=\begin{bmatrix}0 & -H_{11}^{-1}H_{12}\\ -H_{22}^{-1}H_{21} & 0\end{bmatrix}=\begin{bmatrix}-H_{11}^{-1} & 0\\ 0 & -H_{22}^{-1}\end{bmatrix}\begin{bmatrix}0 & H_{12}\\ H_{21} & 0\end{bmatrix}=-D^{-1}N\tag{58}$$

$$b=\begin{bmatrix}-H_{11}^{-1}c_1\\ -H_{22}^{-1}c_2\end{bmatrix}=\begin{bmatrix}-H_{11}^{-1} & 0\\ 0 & -H_{22}^{-1}\end{bmatrix}\begin{bmatrix}c_1\\ c_2\end{bmatrix}=-D^{-1}c\tag{59}$$

$$u^{p+1}=Au^p+b\Leftrightarrow u^{p+1}=-D^{-1}Nu^p+-D^{-1}c\tag{60}$$

Further, some of these decompositions will become useful in the derivations:

$$N^\intercal=\begin{bmatrix}0 & H_{12}\\ H_{21} & 0\end{bmatrix}^\intercal=\begin{bmatrix}0 & H_{21}^\intercal\\ H_{12}^\intercal & 0\end{bmatrix}=\begin{bmatrix}0 & H_{12}\\ H_{21} & 0\end{bmatrix}=N\tag{61}$$

$$H=\begin{bmatrix}H_{11} & H_{12}\\ H_{21} & H_{22}\end{bmatrix}=\begin{bmatrix}H_{11} & 0\\ 0 & H_{22}\end{bmatrix}+\begin{bmatrix}0 & H_{12}\\ H_{21} & 0\end{bmatrix}=D+N\tag{62}$$

$$D^{-1}H=\begin{bmatrix}H_{11}^{-1} & 0\\ 0 & H_{22}^{-1}\end{bmatrix}\begin{bmatrix}H_{11} & H_{12}\\ H_{21} & H_{22}\end{bmatrix}=\begin{bmatrix}I & H_{11}^{-1}H_{12}\\ H_{22}^{-1}H_{21} & I\end{bmatrix}=I+D^{-1}N\tag{63}$$

$$(D^{-1}H)^\intercal=(I+D^{-1}N)^\intercal=I+ND^{-1}\implies HD^{-1}=I+ND^{-1}\tag{64}$$

$$ND^{-1}H=N(I+D^{-1}N)=N+ND^{-1}N=(I+ND^{-1})N=HD^{-1}N\tag{65}$$

4.2.1 Terms Independent of u^p

Starting with the terms independent of u^p, using the fact that $u^*=-H^{-1}c$:

$$\begin{aligned}&-\frac{1}{2}\left(u^{*\intercal}Pu^*\right)\\&=-\frac{1}{2}\left(c^\intercal \cancel{H^{-1}}\cancel{H}D^{-1}(D-N)D^{-1}\cancel{H}\cancel{H^{-1}}c\right)\\&=-\frac{1}{2}\left(c^\intercal D^{-1}c-c^\intercal D^{-1}\underline{N}D^{-1}c\right)\\(62)&=-\frac{1}{2}\left(c^\intercal D^{-1}c-c^\intercal D^{-1}(H-D)D^{-1}c\right)\\&=-\frac{1}{2}\left(2c^\intercal \underline{D^{-1}c}+\underline{c^\intercal D^{-1}}H\underline{D^{-1}c}\right)\\(59)&=\frac{1}{2}b^\intercal Hb+c^\intercal b\quad\text{q.e.d.}\end{aligned}\tag{66}$$

As visible, the equations are a one to one match.

4.2.2 Terms Quadratic in u^p

One can proceed with the terms quadratic in u^p:

$$
\begin{aligned}
& -\frac{1}{2}u^{p\intercal}Pu^p \\
&= -\frac{1}{2}u^{p\intercal}\left(HD^{-1}\tilde{H}D^{-1}H\right)u^p = -\frac{1}{2}u^{p\intercal}\left(HD^{-1}(D-N)D^{-1}H\right)u^p \\
&= -\frac{1}{2}u^{p\intercal}\left(H\underline{D^{-1}H} - HD^{-1}ND^{-1}H\right)u^p \\
\text{(63)} &= -\frac{1}{2}u^{p\intercal}\left(H(I+D^{-1}N) - HD^{-1}ND^{-1}H\right)u^p \\
&= -\frac{1}{2}u^{p\intercal}\left(H + HD^{-1}N - HD^{-1}ND^{-1}H\right)u^p \\
&= -\frac{1}{2}u^{p\intercal}\left(H + HD^{-1}N - HD^{-1}\underline{ND^{-1}H}\right)u^p \\
\text{(65)} &= -\frac{1}{2}u^{p\intercal}\left(H + HD^{-1}N - \underline{HD^{-1}}HD^{-1}N\right)u^p \\
\text{(64)} &= -\frac{1}{2}u^{p\intercal}\left(H + HD^{-1}N - (I+ND^{-1})HD^{-1}N\right)u^p \\
&= -\frac{1}{2}u^{p\intercal}\left(H + \cancel{HD^{-1}N} - \cancel{HD^{-1}N} - ND^{-1}HD^{-1}N\right)u^p \\
&= -\frac{1}{2}u^{p\intercal}\left(H - \underline{ND^{-1}}H\underline{D^{-1}N}\right)u^p \overset{(58)}{=} -\frac{1}{2}u^{p\intercal}\left(H - A^\intercal HA\right)u^p \\
&= \frac{1}{2}u^{p\intercal}\left(A^\intercal HA - H\right)u^p \quad \text{q.e.d.}
\end{aligned}
\tag{67}
$$

Once again, we're able to extract the left hand side from the right hand and can proceed with the derivations.

4.2.3 Terms Linear in u^p

The first point of focus here are the terms of the form $u^{p\intercal}X$

$$
\begin{aligned}
& \frac{1}{2}u^{p\intercal}Pu^* \\
&= -\frac{1}{2}u^{p\intercal}\left(HD^{-1}\tilde{H}D^{-1}\cancel{HH^{-1}}c\right) = -\frac{1}{2}u^{p\intercal}\left(HD^{-1}(D-N)\underline{D^{-1}c}\right) \\
\text{(59)} &= \frac{1}{2}u^{p\intercal}\left(HD^{-1}(D-N)b\right) = \frac{1}{2}u^{p\intercal}\left(H - \underline{HD^{-1}N}\right)b \\
\text{(65)} &= \frac{1}{2}u^{p\intercal}\left(H - \underline{ND^{-1}}H\right)b \overset{(58)}{=} \frac{1}{2}u^{p\intercal}\left(\underline{H} + A^\intercal H\right)b \\
\text{(62)} &= \frac{1}{2}u^{p\intercal}\left(N + D + A^\intercal H\right)b = \frac{1}{2}u^{p\intercal}\left(A^\intercal Hb + (N+D)\underline{b}\right) \\
&= \frac{1}{2}u^{p\intercal}\left(A^\intercal Hb - (N+D)D^{-1}c\right) = \frac{1}{2}u^{p\intercal}\left(A^\intercal Hb\underline{-ND^{-1}}c - c\right) \\
\text{(58)} &= \frac{1}{2}u^{p\intercal}\left(A^\intercal Hb + A^\intercal c - c\right) \\
&= \frac{1}{2}u^{p\intercal}\left(A^\intercal Hb + (A^\intercal - I)c\right) \quad \text{q.e.d.}
\end{aligned}
\tag{68}
$$

The final part of the derivation are the terms in the form Xu^p:

$$
\begin{aligned}
&\frac{1}{2}u^{*\intercal}Pu^p \\
&= -\frac{1}{2}\underline{c^\intercal}\left(\cancel{H^{-1}}\cancel{H}\underline{D^{-1}}\tilde{H}D^{-1}H\right)u^p \overset{(59)}{=} \frac{1}{2}b^\intercal\left((D-N)D^{-1}H\right)u^p \\
(65)\ &= \frac{1}{2}b^\intercal\left(H-\underline{ND^{-1}H}\right)u^p = \frac{1}{2}b^\intercal\left(H-H\underline{D^{-1}N}\right)u^p \\
(58)\ &= \frac{1}{2}b^\intercal\left(H+HA\right)u^p \overset{(62)}{=} \frac{1}{2}b^\intercal\left(N+D+HA\right)u^p \\
&= \frac{1}{2}\left(\underline{b^\intercal}(N+D)+b^\intercal HA\right)u^p \overset{(59)}{=} \frac{1}{2}\left(-c^\intercal D^{-1}(N+D)+b^\intercal HA\right)u^p \\
&= \frac{1}{2}\left(-c^\intercal\underline{D^{-1}N}-c^\intercal+b^\intercal HA\right)u^p \overset{(58)}{=} \frac{1}{2}\left(c^\intercal A-c^\intercal+b^\intercal HA\right)u^p \\
&= \frac{1}{2}\left(c^\intercal(A-I)+b^\intercal HA\right)u^p \quad \text{q.e.d.}
\end{aligned}
\tag{69}
$$

4.2.4 Conclusion

All equations match, and hence it has been successfully shown that:

$$
V(u^{p+1})-V(u^p) = -\frac{1}{2}(u^p-u^*)^\intercal P(u^p-u^*), \quad P = HD^{-1}\tilde{H}D^{-1}H, \quad \tilde{H} = D-N \tag{70}
$$

Thus proving the size of the decrease of the value function.

References

[Johansson et al., 2008] Johansson, B., Keviczky, T., Johansson, M., and Johansson, K. H. (2008). Subgradient methods and consensus algorithms for solving convex optimization problems. In *2008 47th IEEE Conference on Decision and Control*, pages 4185–4190.

[Johansson et al., 2006] Johansson, B., Speranzon, A., Johansson, M., and Johansson, K. H. (2006). Distributed model predictive consensus. In *Proceedings of the 17th Symposium on Mathematical Theory of Networks and Systems*, page 2438–2444.

[Keviczky and Jr., 2021] Keviczky, T. and Jr., M. M. (2021). Lecture slides on networked and distributed control systems [sc42100].

[Olfati-Saber et al., 2007] Olfati-Saber, R., Fax, J. A., and Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233.

[Rawlings and Mayne, 2008] Rawlings, J. and Mayne, D. (2008). *Model Predictive Control: Theory and Design*. Nob Hill Publishing.