Networked and Distributed Control: Assignment 3

Niels Stienen, 5595738

I. Problem 1

Consider the following convex optimization problem with a complicating constraint:

$$\begin{aligned} \underset{x_1, x_2}{\text{minimize}} \quad & f_1(x_1) + f_2(x_2) \\ \text{subject to} \quad & x_1 \in \mathcal{X}_1, \quad x_2 \in \mathcal{X}_2, \\ & h_1(x_1) + h_2(x_2) \leq 0 \end{aligned}$$

where $\mathcal{X}_1, \mathcal{X}_2$ are convex sets and all the functions are convex.
Apply primal decomposition to the problem and show the resulting two subproblems. What is the role of the master problem?

Introducing a coupling variable the system can decomposed into a master problem and two subproblems. The role of the master problem is allocating the amount of resources each subproblem can use through updating the coupling variable.

The first step for primal decomposition is to introduce a coupling variable to split the complicating constraint into two different constraints. Let the coupling variable be y resulting in the problem

$$\begin{aligned} \underset{x_1, x_2}{\text{minimize}} \quad & f_1(x_1) + f_2(x_2) \\ \text{subject to} \quad & x_1 \in \mathcal{X}_1, \quad x_2 \in \mathcal{X}_2, \\ & h_1(x_1) \leq y, \\ & h_2(x_2) \leq -y \end{aligned} \tag{1}$$

Now the problem can be split up resulting in the following master problem:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & \phi_1(y) + \phi_2(y) \\ \text{subject to} \quad & y \in \mathcal{Y} \end{aligned} \tag{2}$$

With the two subproblems:

$$\begin{aligned} \underset{x_1}{\text{minimize}} \quad & f_1(x_1) \\ \text{subject to} \quad & x_1 \in \mathcal{X}_1, \\ & h_1(x_1) \leq y \end{aligned} \tag{3}$$

$$\begin{aligned} \underset{x_2}{\text{minimize}} \quad & f_2(x_2) \\ \text{subject to} \quad & x_2 \in \mathcal{X}_2, \\ & h_2(x_2) \leq -y \end{aligned} \tag{4}$$

Let $p(z)$ be the optimal value of the (possibly non-smooth!) strongly convex optimization problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \mathcal{X}, \\ & h(x) \leq z \end{array}$$

Let λ^* be an optimal dual variable associated with the constraint $h(x) \leq z$. (Keep in mind that the constraint is not necessarily scalar, but can be a vector in general.) Show that $-\lambda^*$ is a subgradient of p at z.

From the dual decomposition of the problem we find

$$d(\lambda) = \inf_{x \in \mathcal{X}} L(x, \lambda) \tag{5}$$

$$= \inf_{x \in \mathcal{X}} f(x) + \lambda^\top (h(x) - z) \tag{6}$$

To show this, we consider the value of $p(\tilde{z})$ where $\tilde{z}$ is not the optimal point based on the work in [1].

$$p(\tilde{z}) = \sup_{\lambda \geq 0} \inf_{x \in \mathcal{X}} \left(f(x) + \lambda^\top (h(x) - \tilde{z}) \right) \tag{7}$$

$$\geq \inf_{x \in \mathcal{X}} \left(f(x) + \lambda^{*\top} (h(x) - \tilde{z}) \right) \tag{8}$$

$$= \inf_{x \in \mathcal{X}} \left(f(x) + \lambda^{*\top} (h(x) - z + z - \tilde{z}) \right) \tag{9}$$

$$= \underbrace{\inf_{x \in \mathcal{X}} \left(f(x) + \lambda^{*\top} (h(x) - z) \right)}_{\text{Problem from Equation 6}} + \lambda^{*\top} (z - \tilde{z}) \tag{10}$$

$$= p(z) - \lambda^{*\top} (\tilde{z} - z) \tag{11}$$

The problem is now decoupled into two parts. One part is the original subproblem to which we know the optimal solution and the other is the only term dependent on $\tilde{z}$. Taking the gradient with respect to $\tilde{z}$ results in

$$\frac{\mathrm{d}p(\tilde{z})}{\mathrm{d}\tilde{z}} = -\lambda^* \tag{12}$$

This holds for all points $\tilde{z}$, thus $-\lambda^*$ is a subgradient of $p(z)$.

II. Problem 2

Consider the combined consensus/projected incremental subgradient method for N agents:

$$x_{k+1}^i = \mathcal{P}_{\mathcal{X}} \left(\sum_{j=1}^{N} \left[W_{ij}^{\phi} \right] \left(x_k^j - \alpha_k g^j (x_k^j) \right) \right), \quad i = 1, ..., N$$

The weight matrix $W \in \mathbb{R}^{N \times N}$ fulfills

$$[W]_{ij} = 0, \text{ if } (i,j) \notin \mathcal{E} \text{ and } i \neq j,$$

$$W = W^\top, \; W \mathbb{1}_N = \mathbb{1}_N, \; \rho \left(W - \frac{\mathbb{1}_N \mathbb{1}_N^\top}{N} \right) \leq \gamma < 1,$$

where $\rho(\cdot)$ is the spectral radius and $\mathbb{1}_N \in \mathbb{R}^N$ is the column vector with all elements equal to one. Show that as $\phi \to \infty$ the combined consensus/projected incremental subgradient method becomes a standard subgradient method.

Writing out the projected incremental subgradient we get:

$$\begin{bmatrix} x_{k+1}^1 \\ x_{k+1}^2 \\ \vdots \\ x_{k+1}^N \end{bmatrix} = \mathcal{P}_{\mathcal{X}} \begin{bmatrix} W_{11}^{\phi}(x_k^1 - \alpha_k g^1(x_k^1)) + W_{12}^{\phi}(x_k^2 - \alpha_k g^2(x_k^2)) + \ldots + W_{2N}^{\phi}\big(x_k^N - \alpha_k g^N(x_k^N)\big) \\ W_{21}^{\phi}(x_k^1 - \alpha_k g^1(x_k^1)) + W_{22}^{\phi}(x_k^2 - \alpha_k g^2(x_k^2)) + \ldots + W_{1N}^{\phi}\big(x_k^N - \alpha_k g^N(x_k^N)\big) \\ \vdots \\ W_{N1}^{\phi}(x_k^1 - \alpha_k g^1(x_k^1)) + W_{N2}^{\phi}(x_k^2 - \alpha_k g^2(x_k^2)) + \ldots + W_{NN}^{\phi}\big(x_k^N - \alpha_k g^N(x_k^N)\big) \end{bmatrix} \tag{13}$$

This equation can be written in a compact form by using vector notation

$$\mathbf{x}_{k+1} = W^{\phi}\mathcal{P}_{\mathcal{X}}(\mathbf{x}_k - \alpha_k \mathbf{g}(\mathbf{x}_k)) \tag{14}$$

Then if and only if $\lim_{\phi\to\infty} W^{\phi} = \mathbb{1}^{\top}\mathbb{1}/N$ this equation is equivalent to the form

$$x_{k+1} = \mathcal{P}_{\mathcal{X}}\left(x_k - \alpha_k \sum_{i=1}^{N} g_k^i\right) \tag{15}$$

To show that $\lim_{\phi\to\infty} W^{\phi} = \mathbb{1}^{\top}\mathbb{1}/N$ is indeed true theorem 1 from ”Fast linear iterations for distributed averaging” [2] is used

Theorem 1:

$$\lim_{\phi\to\infty} W^{\phi} = \frac{\mathbb{1}^{\top}\mathbb{1}}{N} \tag{16}$$

holds, if and only if

$$\mathbb{1}^{\top}W = \mathbb{1}^{\top}, \tag{17}$$

$$W\mathbb{1} = \mathbb{1}, \tag{18}$$

$$\rho\left(W - \frac{\mathbb{1}\mathbb{1}^{\top}}{N}\right) < 1 \tag{19}$$

The conditions stated in Eq. 18, 19 are given in the problem which leaves only Eq. 17 to show. Making use of the fact that $W = W^{\top}$ it is trivial to show

$$\mathbb{1}^{\top}W = \mathbb{1}^{\top} \tag{20}$$

$$\left(\mathbb{1}^{\top}W\right)^{\top} = \mathbb{1} \tag{21}$$

$$W^{\top}\mathbb{1} = \mathbb{1} \tag{22}$$

$$W\mathbb{1} = \mathbb{1} \tag{23}$$

Which is given, thus in the limit W^{ϕ} converges giving the following equivalence

$$\mathbf{x}_{k+1} = \frac{\mathbb{1}^{\top}\mathbb{1}}{N}\mathcal{P}_{\mathcal{X}}(\mathbf{x}_k - \alpha_k \mathbf{g}(\mathbf{x}_k)) \Leftrightarrow \mathbf{x}_{k+1} = \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}_k - \alpha_k \sum_{i=1}^{N} g_k^i\right) \tag{24}$$

Intuitively, this is also what one would expect given that W is doubly stochastic. A doubly stochastic matrix is a square matrix of nonnegative real numbers for which each of its rows and columns sums to 1:

$$\sum_i W_{ij} = \sum_j W_{ij} = 1 \tag{25}$$

Therefore W can be seen as an averaging matrix which due to the property of the associated communication graph $\mathcal{G}$ being strongly connected will converge as defined in Equation 16.

III. Problem 3

Consider the positive definite quadratic function partitioned into two sets of variables

$$V(u) = \frac{1}{2}u^\top H u + c^\top u + d \tag{26}$$

$$V(u_1, u_2) = \frac{1}{2}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}^\top \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}^\top \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + d \tag{27}$$

in which $H > 0$ meaning that it is symmetric positive definite.

Given an initial point (u_1^p, u_2^p), show that the next iteration is:

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

To show that the next iteration is indeed as given, the partial derivative of the cost function with respect to u_1, u_2 is set to zero. Making use of the symmetry of H the derivative can be further simplified by combining terms.

$$\left.\frac{\partial V(u_1, u_2)}{\partial u_1}\right|_{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} u_1^{p+1} \\ u_2^p \end{bmatrix}} = \frac{1}{2}H_{11}u_1 + \frac{1}{2}H_{11}^\top u_1 + \frac{1}{2}H_{12}u_2 + \frac{1}{2}\left(u_2^\top H_{21}\right)^\top + c_1 = 0 \tag{28}$$

$$H_{11}u_1^{p+1} + H_{12}u_2^p + c_1 = 0 \tag{29}$$

$$-H_{11}u_1^{p+1} = H_{12}u_2^p + c_1 \tag{30}$$

$$u_1^{p+1} = -H_{11}^{-1}(H_{12}u_2^p + c_1) \tag{31}$$

The same steps are used to arrive at the update equation for u_2

$$\left.\frac{\partial V(u_1, u_2)}{\partial u_2}\right|_{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} u_1^{p} \\ u_2^{p+1} \end{bmatrix}} = \frac{1}{2}H_{22}u_2 + \frac{1}{2}H_{22}^\top u_2 + \frac{1}{2}H_{21}u_1 + \frac{1}{2}\left(u_1^\top H_{12}\right)^\top + c_2 = 0 \tag{32}$$

$$H_{22}u_2^{p+1} + H_{21}u_1^p + c_2 = 0 \tag{33}$$

$$-H_{22}u_2^{p+1} = H_{21}u_1^p + c_2 \tag{34}$$

$$u_2^{p+1} = -H_{22}^{-1}(H_{21}u_1^p + c_2) \tag{35}$$

Combining these two update steps into one iteration update the update procedure is found as

$$\begin{aligned} u_1^{p+1} &= -H_{11}^{-1}(H_{12}u_2^p + c_1) \\ u_2^{p+1} &= -H_{22}^{-1}(H_{21}u_1^p + c_2) \end{aligned} \tag{36}$$

$$\begin{bmatrix} u_1^{p+1} \\ u_2^{p+1} \end{bmatrix} = \begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix}\begin{bmatrix} u_1^p \\ u_2^p \end{bmatrix} + \begin{bmatrix} -H_{11}^{-1}c_1 \\ -H_{22}^{-1}c_2 \end{bmatrix} \tag{37}$$

The procedure can be summarized as $u^{p+1} = Au^p + b$ in which the iteration matrix A and constant b are given by

$$A = \begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix}, \quad b = \begin{bmatrix} -H_{11}^{-1}c_1 \\ -H_{22}^{-1}c_2 \end{bmatrix}$$

Establish that the optimization procedure converges by showing the iteration matrix is stable

$$|eig(A)| < 1$$

The Householder-John theorem [3] is used to show that the iteration matrix is stable. The theorem states:

Theorem 2: (The Householder-John theorem) [3]
If C and D are real matrices such that both C and $C - D - D^\top$ are symmetric positive definite, then the spectral radius of $F = -(C - D)^{-1}D$ is strictly less than one.

Setting $F = A$ the matrices C and D can be found and checked whether they fulfill the condition stated.

$$\underbrace{\begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix}}_{F} = -\left(\underbrace{\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}}_{C} - \underbrace{\begin{bmatrix} 0 & H_{12} \\ H_{21} & 0 \end{bmatrix}}_{D}\right)^{-1} \underbrace{\begin{bmatrix} 0 & H_{12} \\ H_{21} & 0 \end{bmatrix}}_{D} \tag{38}$$

$$= \begin{bmatrix} H_{11} & 0 \\ 0 & H_{22} \end{bmatrix}^{-1} \begin{bmatrix} 0 & H_{12} \\ H_{21} & 0 \end{bmatrix} \tag{39}$$

$$= \begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix} \tag{40}$$

The matrix C is the same as the original matrix H which is symmetric positive definite. Due to H being symmetric the following equivalence is used $H_{12} = H_{21}^\top$ resulting in

$$C - D - D^\top = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} - \begin{bmatrix} 0 & H_{12} \\ H_{21} & 0 \end{bmatrix} - \begin{bmatrix} 0 & H_{21}^\top \\ H_{12}^\top & 0 \end{bmatrix} = \begin{bmatrix} H_{11} & -H_{12} \\ -H_{21} & H_{22} \end{bmatrix} \tag{41}$$

The matrix $C - D - D^\top = \bar{H}$ which is also given to be symmetric positive definite in the assignment. It has been shown that both of the conditions hold. Hence, by application of Householder-John the matrix A has a spectral radius of strictly less than one. The spectral radius returns the largest eigenvalue of a matrix thus the iteration matrix A must be stable as

$$\rho(A) = \max\left(\|\lambda\|_1\right) \geq \lambda_i, \quad \forall i \tag{42}$$

Given that the iteration converges, show that it produces the same solution as $u^* = -H^{-1}c$

Taking the limit of the update sequence the sequence converges to

$$\lim_{p\to\infty} u^{p+1} = Au^p + b \implies u^\infty = Au^\infty + b \tag{43}$$

Rewriting the above limit to separate the u^∞ one gets

$$u^\infty = Au^\infty + b \tag{44}$$
$$(I - A)u^\infty = b \tag{45}$$
$$u^\infty = (I - A)^{-1}b \tag{46}$$

Since $u^* = u^\infty$ it needs to be shown that

$$-H^{-1}c = (I - A)^{-1}b \tag{47}$$
$$-H^{-1}c = \left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix}\right)^{-1} \begin{bmatrix} -H_{11}^{-1}c_1 \\ -H_{22}^{-1}c_2 \end{bmatrix} \tag{48}$$
$$-H^{-1}c = -\begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix}^{-1} \begin{bmatrix} H_{11}^{-1} & 0 \\ 0 & H_{22}^{-1} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{49}$$
$$H^{-1}c = \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix}^{-1} \begin{bmatrix} H_{11}^{-1} & 0 \\ 0 & H_{22}^{-1} \end{bmatrix} c \tag{50}$$
$$\begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix} H^{-1}c = \begin{bmatrix} H_{11}^{-1} & 0 \\ 0 & H_{22}^{-1} \end{bmatrix} c \tag{51}$$
$$\begin{bmatrix} H_{11} & 0 \\ 0 & H_{22} \end{bmatrix} \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix} H^{-1}c = c \tag{52}$$
$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} H^{-1}c = c \tag{53}$$
$$HH^{-1}c = c \tag{54}$$
$$Ic = c \tag{55}$$
$$c = c \tag{56}$$

Thus the iteration procedure produces the same solution as $u^* = -H^{-1}c$.

IV. Problem 4

Consider again the iteration defined in the previous problem. Prove that the cost function is monotonically decreasing when optimizing one variable at a time

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

and show that the following expression gives the size of the decrease

$$V\left(u^{p+1}\right) - V(u^p) = -\frac{1}{2}(u^p - u^*)^\top P(u^p - u^*)$$

in which

$$P = HD^{-1}\tilde{H}D^{-1}H, \quad \tilde{H} = D - N,$$
$$D = \begin{bmatrix} H_{11} & 0 \\ 0 & H_{22} \end{bmatrix}, \quad N = \begin{bmatrix} 0 & H_{12} \\ H_{21} & 0 \end{bmatrix}$$

and $u^* = -H^{-1}c$ is the optimum.

To simplify later steps first a coordinate transformation is used

$$z^p = u^p - u^* \quad \rightarrow \quad u^p = z^p + u^* \tag{57}$$

For the next step u^{p+1} the coordinate transform is then

$$z^{p+1} = u^{p+1} - u^* \tag{58}$$
$$= Au^p + b - u^* \tag{59}$$
$$= A(z^p + u^*) + b - u^* \tag{60}$$
$$= Az^p + \underbrace{Au^* + b}_{=u^*} - u^* \tag{61}$$
$$z^{p+1} = Az^p \tag{62}$$

Since now both the current and future step can be described in terms of z^p the superscript p is dropped in the later equations.

Applying the same transformation to the cost function $V(u)$ the cost function can be rewritten in terms of z by making use of $u^* = -H^{-1}c$ as

$$V(u) = \frac{1}{2}(z + u^*)^\top H(z + u^*) + c^\top(z + u^*) + d \tag{63}$$
$$= \frac{1}{2}z^\top Hz + \frac{1}{2}z^\top Hu^* + \frac{1}{2}u^{*\top}Hz + \frac{1}{2}u^{*\top}Hu^* + c^\top z + c^\top u^* + d \tag{64}$$
$$= \frac{1}{2}z^\top Hz - \frac{1}{2}z^\top HH^{-1}c - \frac{1}{2}c^\top H^{-1}Hz + \frac{1}{2}c^\top H^{-1}HH^{-1}c + c^\top z - c^\top H^{-1}c + d \tag{65}$$
$$= \frac{1}{2}z^\top Hz - \frac{1}{2}z^\top c - \frac{1}{2}c^\top z + \frac{1}{2}c^\top H^{-1}c + c^\top z - c^\top H^{-1}c + d \tag{66}$$
$$V(u) = \frac{1}{2}z^\top Hz \underbrace{- \frac{1}{2}c^\top H^{-1}c + d}_{\text{Constant}} \tag{67}$$

After the coordinate transformation, the cost function is reduced to one quadratic term plus a constant which can be left out as for both parts that need to be proven the constant term will immediately cancel out. First, the size of the decrease will be shown by writing out the difference equation

$$V\big(z^{p+1}\big) - V(z^p) = -\frac{1}{2}z^\top P z \tag{68}$$

$$V(Az^p) - V(z^p) = \tag{69}$$

$$\frac{1}{2}z^\top A^\top H A z - \frac{1}{2}z^\top H z = \tag{70}$$

$$\frac{1}{2}z^\top\big(A^\top H A - H\big)z = -\frac{1}{2}z^\top P z \tag{71}$$

$$\frac{1}{2}z^\top\big(H - A^\top H A\big)z = \frac{1}{2}z^\top P z \tag{72}$$

$$\tag{73}$$

Thus it needs to be shown that

$$H - A^\top H A = P \tag{74}$$

This is done through writing out the entire matrix P and finding the corresponding terms to reduce it back to the left-hand side of the expression.

$$P = HD^{-1}\tilde{H}D^{-1}H \tag{75}$$

$$= HD^{-1}(D-N)D^{-1}H \tag{76}$$

$$= \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} H_{11}^{-1} & 0 \\ 0 & H_{22}^{-1} \end{bmatrix} \begin{bmatrix} H_{11} & -H_{12} \\ -H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} H_{11}^{-1} & 0 \\ 0 & H_{22}^{-1} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{77}$$

$$= \begin{bmatrix} I & H_{12}H_{22}^{-1} \\ H_{21}H_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} H_{11} & -H_{12} \\ -H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix} \tag{78}$$

$$= \begin{bmatrix} H_{11} - H_{12}H_{22}^{-1}H_{21} & -H_{12} + H_{12}H_{22}^{-1}H_{22} \\ -H_{21} + H_{21}H_{11}^{-1}H_{11} & H_{22} - H_{21}H_{11}^{-1}H_{12} \end{bmatrix} \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix} \tag{79}$$

$$= \begin{bmatrix} H_{11} - H_{12}H_{22}^{-1}H_{21} & 0 \\ 0 & H_{22} - H_{21}H_{11}^{-1}H_{12} \end{bmatrix} \begin{bmatrix} I & H_{11}^{-1}H_{12} \\ H_{22}^{-1}H_{21} & I \end{bmatrix} \tag{80}$$

$$= \begin{bmatrix} H_{11} - H_{12}H_{22}^{-1}H_{21} & H_{12} - H_{12}H_{22}^{-1}H_{21}H_{11}^{-1}H_{12} \\ H_{21} - H_{21}H_{11}^{-1}H_{12}H_{22}^{-1}H_{21} & H_{22} - H_{21}H_{11}^{-1}H_{12} \end{bmatrix} \tag{81}$$

$$= \underbrace{\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}}_{=H} + \begin{bmatrix} -H_{12}H_{22}^{-1}H_{21} & -H_{12}H_{22}^{-1}H_{21}H_{11}^{-1}H_{12} \\ -H_{21}H_{11}^{-1}H_{12}H_{22}^{-1}H_{21} & -H_{21}H_{11}^{-1}H_{12} \end{bmatrix} \tag{82}$$

$$= H + \begin{bmatrix} H_{12}H_{22}^{-1}H_{21} & H_{12} \\ H_{21} & H_{21}H_{11}^{-1}H_{12} \end{bmatrix} \underbrace{\begin{bmatrix} 0 & -H_{11}^{-1}H_{12} \\ -H_{22}^{-1}H_{21} & 0 \end{bmatrix}}_{=A} \tag{83}$$

$$= H - \begin{bmatrix} -H_{12}H_{22}^{-1}H_{21} & -H_{12} \\ -H_{21} & -H_{21}H_{11}^{-1}H_{12} \end{bmatrix} A \tag{84}$$

$$= H - \underbrace{\begin{bmatrix} 0 & -H_{12}H_{22}^{-1} \\ -H_{21}H_{11}^{-1} & 0 \end{bmatrix}}_{=A^\top} \underbrace{\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}}_{=H} A \tag{85}$$

$$P = H - A^\top H A \tag{86}$$

And thus it is shown that the size of the decrease is indeed

$$V\big(z^{p+1}\big) - V(z^p) = -\frac{1}{2}z^\top P z \tag{87}$$

Next it needs to be shown that the cost function is monotonically decreasing when optimizing one variable at a time, for this we fill in the cost function on both sides, where again the constant term cancels out and is thus left out in the equation.

$$V\left(z^{p+1}\right) < V(z^p) \tag{88}$$

$$\frac{1}{2}z^\top A^\top HAz < \frac{1}{2}z^\top Hz \tag{89}$$

$$A^\top HA < H \tag{90}$$

Referring back to problem 3b it was shown that $\rho(A) < 1$. Therefore $A^\top HA < H$ holds and the cost function is indeed monotonically decreasing when optimizing one variable at a time.

References

[1] S. Boyd, "Ee364b: Convex optimization ii, lecture notes: Decomposition," https://stanford.edu/class/ee364b/lectures/decomposition_notes.pdf, 2021.

[2] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," in *42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475)*, vol. 5, 2003, pp. 4997–5002 Vol.5.

[3] A. Hansen, "Mathematical tripos part ii: Michaelmas term 2020, numerical analysis – lecture 17," 2020, lecture notes. [Online]. Available: https://www.damtp.cam.ac.uk/research/afha/lectures/Part_II_NumAn/