Design of Optimal Sparse Feedback Gains via the Alternating Direction Method of Multipliers

Fu Lin, Makan Fardad, and Mihailo R. Jovanović

Jineng Ren
ECE Department, University of Minnesota

Class Report, March 9, 2017
Consider the following control problem

\[
\dot{x} = Ax + B_1d + B_2u \\
z = Cx + Du \\
u = -Fx
\]  
(1.1)
Consider the following control problem

\[ \dot{x} = Ax + B_1d + B_2u \]
\[ z = Cx + Du \]
\[ u = -Fx \] (1.1)

- \( d \) and \( u \) are the disturbance and control inputs
- \( z \) is the performance output
- \( C \) and \( D \) contain the state performance weight \( Q \) and the control performance weight \( P \)
- matrix \( F \) is the state feedback gain
Search for $F \in S$ that minimizes the $\mathcal{H}_2$ norm of the transfer function from $d$ to $z$. 

Solve the optimization problem

$$\min_{F \in S} J(F)$$

subject to $F \in S$, (1.2)

where $J(F) = \begin{cases} \text{trace}(B^T P(F) B), & F \text{ stabilizing} \\ \infty, & \text{otherwise} \end{cases}$, (1.3)

$P(F)$ denotes the closed-loop observability Gramian.
Search for $F \in S$ that minimizes the $\mathcal{H}_2$ norm of the transfer function from $d$ to $z$. Solve the optimization problem

$$\begin{align*}
\text{minimize} & \quad J(F) \\
\text{subject to} & \quad F \in S,
\end{align*}$$

(1.2)

where

$$J(F) = \begin{cases} 
\text{trace}(B_1^TP(F)B_1), & F \text{ stabilizing} \\
\infty, & \text{otherwise}
\end{cases}$$

(1.3)

$P(F)$ denotes the closed-loop observability Gramian.
Basic idea:

- **Step 1:** Identify sparsity patterns of feedback gains; incorporating sparsity-promoting penalty functions; the added terms penalize the number of communication links in the distributed controller.
Basic idea:

- Step 1: Identify sparsity patterns of feedback gains; incorporating sparsity-promoting penalty functions; the added terms penalize the number of communication links in the distributed controller.

- Step 2: Optimize feedback gains subject to structural constraints determined by the identified sparsity patterns obtained above.
Adding the penalty term

\[
\text{minimize} \quad J(F) + \gamma g(F) \quad (1.4)
\]

where the penalty function \( g(F) \) can be

(1) \( g_0(F) = \text{card}(F); \)
Adding the penalty term

\[
\text{minimize} \quad J(F) + \gamma g(F) \quad (1.4)
\]

where the penalty function \( g(F) \) can be

1. \( g_0(F) = \text{card}(F); \)
2. \( g_1(F) = \| F \|_1 = \sum_{i,j} |F|_{ij}; \)
Adding the penalty term

\[
\text{minimize} \quad J(F) + \gamma g(F) \tag{1.4}
\]

where the penalty function \( g(F) \) can be

1. \( g_0(F) = \text{card}(F) \);
2. \( g_1(F) = \| F \|_1 = \sum_{i,j} |F|_{ij} \);
3. \( g_2(F) = \sum_{i,j} W_{i,j} |F|_{ij} \).
Adding the penalty term

\[
\text{minimize} \quad J(F) + \gamma g(F) \quad (1.4)
\]

where the penalty function \( g(F) \) can be

(1) \( g_0(F) = \text{card}(F); \)
(2) \( g_1(F) = \| F \|_1 = \sum_{i,j} |F|_{ij}; \)
(3) \( g_2(F) = \sum_{i,j} W_{i,j} |F|_{ij}. \)
(4) sum-of-logs function

\[
g_3(F) = \sum_{i,j} \log(1 + |F|_{ij}/\epsilon) \quad (1.5)
\]
In block sparse design, $g$ is determined by

- $\sum_{i,j} W_{i,j} \| G_{ij} \|_F$;
- $\sum_{i,j} \text{card}(\| G_{ij} \|_F)$;
- $\sum_{i,j} \log(1 + \| G_{ij} \|_F / \epsilon)$. 
Consider the following constrained optimization problem

\[
\begin{align*}
\text{minimize} & \quad J(F) + \gamma g(G) \\
\text{subject to} & \quad F - G = 0.
\end{align*}
\] (2.6)
Consider the following constrained optimization problem

$$\text{minimize} \quad J(F) + \gamma g(G)$$

subject to \( F - G = 0. \) \hspace{1cm} (2.6)

The augmented Lagrangian associated with the constrained problem

$$\mathcal{L}_\rho(F, F, \Lambda) = J(F) + \gamma g(G) + \text{trace}(\Lambda^T(F - G)) + \rho/2\|F - G\|^2_F$$

where \( \Lambda \) is the dual variable (i.e., the Lagrange multiplier), \( \rho \) is a positive scalar.
Consider the following constrained optimization problem

\[
\begin{align*}
\text{minimize} & \quad J(F) + \gamma g(G) \\
\text{subject to} & \quad F - G = 0. \tag{2.6}
\end{align*}
\]

The augmented Lagrangian associated with the constrained problem

\[
\mathcal{L}_\rho(F, F, \Lambda) = J(F) + \gamma g(G) + \text{trace}(\Lambda^T(F - G)) + \rho/2\|F - G\|_F^2
\]

where \(\Lambda\) is the dual variable (i.e., the Lagrange multiplier), \(\rho\) is a positive scalar.

Alternating update:

\[
\begin{align*}
F^{k+1} & := \arg\min_F \mathcal{L}_\rho(F, G^k, \Lambda^k) \\
G^{k+1} & := \arg\min_G \mathcal{L}_\rho(F^{k+1}, G, \Lambda^k) \\
\Lambda^{k+1} & := \Lambda^k + \rho(F^{k+1} - G^{k+1}).
\end{align*}
\]
To update $G$, we need to solve the following problem

$$\text{minimize} \quad \gamma g(G) + \left(\frac{\rho}{2}\right) \|G - V^k\|$$

(2.7)

1) Weighted $l_1$ norm: The unique solution to (2.7)

**Soft thresholding**

$$G_{ij}^* = \begin{cases} 
(1 - \frac{a}{|V_{ij}|})V_{ij}, & |V_{ij}| > a \\
0, & |V_{ij}| > a
\end{cases}$$

where $a = (\gamma/\rho)W_{ij}$. 
2) Cardinality function: The unique solution to (2.7)

\[ G_{ij}^* = \begin{cases} 
|V_{ij}|, & |V_{ij}| > b \\
0, & |V_{ij}| \leq b,
\end{cases} \]

where \( b = \sqrt{2\gamma/\rho} \).
3) Sum-of-logs function (1.5): As shown in (F. Lin, 2012), the unique solution to (2.7)

\[
G^*_{ij} = \begin{cases} 
0, & \Delta \leq 0 \text{ or } \{\Delta > 0 \text{ and } r^+\} \\
 r^+ V_{ij}, & \Delta > 0 \text{ and } r^- \leq 0 \text{ and } 0 < r^+ \leq 1 \\
 G^0, & \Delta > 0 \text{ and } 0 \leq r^\pm 
\end{cases}
\]

where

\[
\Delta = (|V_{ij}| + \epsilon)^2 - 4(\gamma/\rho)
\]

\[
r^\pm = (|V_{ij}| - \epsilon \pm \sqrt{\Delta})/(2|V_{ij}|)
\]

and \( G^0 := \operatorname{argmin}\{\phi_{ij}(r^+ V_{ij}), \phi_{ij}(0)\} \).
minimize \( \phi(F) = \gamma J(F) + (\rho/2)\|F - U^k\| \)

Setting \( \nabla \phi := \nabla J + \rho(F - U^k) \) to zero yields the necessary conditions for optimality

\[
2(RF - B_2^T P)L + \rho(F - U^k) = 0;
\]
where \( L \) and \( P \) are determined by

\[
(A - B_2F)L + L(A - B_2F)^T = -B_1B_1^T
\]

\[
(A - B_2F)^T P + P(A - B_2F) = -(Q + F^T RF).
\]
Step 2: Use descent algorithms (e.g., Newton's method) to solve (1.2).

- Given an initial gain $F \in S$
- updating $F$ according to $F^{i+1} = F^i + s^i \tilde{F}^i$; here, $s^i$ is the step-size and $\tilde{F}^i \in S$ is the Newton direction that is determined by the minimizer of the second-order approximation of the objective function (1.3).
Problem Formulation

Algorithm

Experiments

end

Mass-spring system

The state-space representation is given by (1.1) with

\[
A = \begin{bmatrix} O & I \\ T & O \end{bmatrix}; \quad B_1 = B_2 = \begin{bmatrix} O \\ I \end{bmatrix}
\]

\(T\): tridiagonal Toeplitz matrix with -2 on its main diagonal and 1 on its first sub- and super-diagonal.

Fig. 1: Sparsity patterns of \(F^\star = [F^\star_p \ F^\star_v] \in \mathbb{R}^{50 \times 100}\) for the mass-spring system obtained using weighted \(\ell_1\) norm with (a) \(\gamma = 10^{-4}\) and (b) \(\gamma = 0.0105\). As \(\gamma\) increases, the number of nonzero sub- and super-diagonals of \(F^\star_p\) and \(F^\star_v\) decreases. The diagonals of (c) \(F^\star_p\) and (d) \(F^\star_v\) for different values of \(\gamma\): \(10^{-4}\) (○), 0.0281 (+), and 0.1 (*). The diagonals of the centralized position and velocity gains are almost identical to (○).
Mass-spring system

Fig. 1: Sparsity patterns of $F^\star = \begin{bmatrix} F^\star_p & F^\star_v \end{bmatrix} \in \mathbb{R}^{50 \times 100}$ for the mass-spring system obtained using weighted $\ell_1$ norm with (a) $\gamma = 10^{-4}$ and (b) $\gamma = 0.0105$. As $\gamma$ increases, the number of nonzero sub- and super-diagonals of $F^\star_p$ and $F^\star_v$ decreases. The diagonals of (c) $F^\star_p$ and (d) $F^\star_v$ for different values of $\gamma$: $10^{-4}$ ($\circ$), $0.0281$ ($+$), and $0.1$ ($\ast$). The diagonals of the centralized position and velocity gains are almost identical to ($\circ$).

(a) $\text{card}(F^\star)/\text{card}(F_c)$
(b) $(J(F^\star) - J(F_c))/J(F_c)$
(c) $\gamma$ 0.01 0.04 0.10
$\text{card}(F^\star)/\text{card}(F_c)$ 9.4% 5.8% 2.0%
$(J(F^\star) - J(F_c))/J(F_c)$ 0.8% 2.3% 7.8%

Fig. 2: (a) The sparsity level and (b) the performance degradation of $F^\star$ compared to the centralized gain $F_c$ for mass-spring system. (c) Sparsity vs. performance: using 2% of nonzero elements, $\mathcal{H}_2$ performance of $F^\star$ is only 7.8% worse than performance of $F_c$. 

University of Minnesota
Network with 100 unstable nodes

\[
\begin{bmatrix}
\dot{x}_{1i} \\
\dot{x}_{2i}
\end{bmatrix} = \begin{bmatrix}
1 & 1 \\
1 & 2
\end{bmatrix} \begin{bmatrix}
x_{1i} \\
x_{2i}
\end{bmatrix} + \sum_{j \neq i} e^{-\alpha(i,j)} \begin{bmatrix}
x_{1j} \\
x_{2j}
\end{bmatrix} + \begin{bmatrix}
0 \\
1
\end{bmatrix} (d_i + u_i)
\]
Network with 100 unstable nodes

\[
\begin{bmatrix}
\dot{x}_{1i} \\
\dot{x}_{2i}
\end{bmatrix} =
\begin{bmatrix}
1 & 1 \\
1 & 2
\end{bmatrix}
\begin{bmatrix}
x_{1i} \\
x_{2i}
\end{bmatrix} + 
\sum_{j \neq i} e^{-\alpha(i,j)}
\begin{bmatrix}
x_{1j} \\
x_{2j}
\end{bmatrix} +
\begin{bmatrix}
0 \\
1
\end{bmatrix} (d_i + u_i)
\]

Fig. 3: (a)-(c) The localized communication graphs of distributed controllers obtained by solving (SP) for different values of $\gamma$ for the network with 100 nodes. Note that the communication graph does not have to be connected since the nodes are dynamically coupled to each other and allowed to measure their own states. (d) The optimal trade-off curve between the $H_2$ performance degradation and the sparsity level of $F^*$ compared to the centralized gain $F_c$. 

Problem Formulation Algorithm Experiments

Block sparsity: bio-chemical reaction

\[ \dot{x}_i = [A]_{ii} x_i - \frac{1}{2} \sum_{j=1}^{N} (i - j)(x_i - x_j) + [B_1]_{ii} d_i + [B_2]_{ii} u_i \]
Block sparsity: bio-chemical reaction

\[ \dot{x}_i = [A]_{ii} x_i - \frac{1}{2} \sum_{j=1}^{N} (i - j)(x_i - x_j) + [B_1]_{ii} d_i + [B_2]_{ii} u_i \]

Fig. 4: The sparse feedback gains obtained by solving (SP) using (a) the weighted sum of Frobenius norms with $\gamma = 3.6$ and (b) the weighted $\ell_1$ norm (9) with $\gamma = 1.3$. Here, $F \in \mathbb{R}^{5 \times 15}$ is partitioned into 25 blocks $F_{ij} \in \mathbb{R}^{1 \times 3}$. Both feedback gains have the same number of nonzero elements (indicated by dots) and close $H_2$ performance (less than 1% difference), but different number of nonzero blocks (indicated by boxes). Communication graphs of (c) the block sparse feedback gain in (a), and (d) the sparse feedback gain in (b) (red color highlights the additional links). An arrow pointing from node $i$ to node $j$ indicates that $i$ uses measurements from $j$. 

V. CONCLUDING REMARKS

We design sparse and block sparse state feedback gains that optimize the $H_2$ performance of distributed systems. The design procedure consists of a structure identification step and a “polishing” step. In the identification step, we employ the ADMM algorithm to solve the sparsity-promoting optimal control problem, whose solution gradually moves from the centralized gain to...
Thanks!

renxx282@umn.edu