Optimal sensor and actuator selection in distributed systems

Mihailo Jovanović
ee.usc.edu/mihailo

joint work with

Armin Zare
Neil Dhingra

IMA Sensor Location Workshop
Motivating applications

networks of dynamical systems

flexible wing aircraft

- **Challenge:** sensor/actuator placement
Context

**Rich History**

- distributed parameter systems literature

  *John Burns’ talk yesterday: outstanding overview!*

**Lessons Learned**

- importance of **problem formulation**
  well-posedness; selection: context dependent

- optimal estimation/control
  much better tool for selection than observability/controllability

- **difficult to solve:** nonconvex, computationally challenging
Context

- Rich history
  - distributed parameter systems literature
  
  *John Burns’ talk yesterday: outstanding overview!*

- Lessons learned
  - importance of problem formulation
    - well-posedness; selection: context dependent
  
  - optimal estimation/control
    - much better tool for selection than observability/controllability
  
  - difficult to solve: nonconvex, computationally challenging

- Why now?
  - applications: networks, distributed sensor/actuator arrays
  
  - optimization: tremendous advances during the last decade
**OBJECTIVE**

select a subset of available sensors/actuators

that provides

“acceptable” degradation of estimation/control quality
Selection via regularization

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

\[\downarrow \quad \downarrow\]

estimation/control \quad proxy for

quality \quad selection

\[\gamma > 0 \quad \rightarrow \quad \text{performance vs “complexity” tradeoff}\]
TRADE-OFF CURVE

* performance vs “complexity”
Minimum variance control problem

dynamics: \( \dot{x} = A x + B_1 d + B_2 u \)

objective function: \( J = \lim_{t \to \infty} \mathbb{E} \left( x^T(t) Q x(t) + u^T(t) R u(t) \right) \)

memoryless controller: \( u = -F x \)
Minimum variance control problem

dynamics: \[ \dot{x} = Ax + B_1 d + B_2 u \]

objective function: \[ J = \lim_{t \to \infty} E \left( x^T(t) Q x(t) + u^T(t) R u(t) \right) \]

memoryless controller: \[ u = -F x \]

- **CLOSED-LOOP VARIANCE AMPLIFICATION**

\[ J \text{ – non-convex function of } F \]
No structural constraints

- **SDP CHARACTERIZATION**

minimize \( X, F \) \[ \text{trace} \left( (Q + F^T R F) X \right) \]

subject to \( (A - B_2 F) X + X (A - B_2 F)^T + B_1 B_1^T = 0 \)

\( X \succ 0 \)
No structural constraints

- **SDP CHARACTERIZATION**

\[
\begin{align*}
\text{minimize} \quad & \quad \text{trace} \left( (Q + F^T R F) X \right) \\
\text{subject to} \quad & \quad (A - B_2 F) X + X (A - B_2 F)^T + B_1 B_1^T = 0 \\
& \quad X \succ 0
\end{align*}
\]

\[
\begin{align*}
\quad & \quad \star \quad \text{change of variables:} \quad F X = Y
\end{align*}
\]

\[
\begin{align*}
\text{minimize} \quad & \quad \text{trace} \left( Q X \right) + \text{trace} \left( R Y X^{-1} Y^T \right) \\
\text{subject to} \quad & \quad (A X - B_2 Y) + (A X - B_2 Y)^T + B_1 B_1^T = 0 \\
& \quad X \succ 0
\end{align*}
\]

Schur complement \(\Rightarrow\) **SDP characterization**
**Riccati-based-characterization**

globally optimal controller

\[
A^T P + PA - PB_2 R^{-1} B_2^T P + Q = 0 \\
F_c = R^{-1} B_2^T P
\]
**STRUCTURAL CONSTRAINTS**  \( F \in S \)

<table>
<thead>
<tr>
<th>centralized</th>
<th>fully-decentralized</th>
<th>localized</th>
</tr>
</thead>
</table>
| \[
\begin{bmatrix}
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
* & * & * & *
\end{bmatrix}
\] | \[
\begin{bmatrix}
* \\
* \\
* \\
*
\end{bmatrix}
\] | \[
\begin{bmatrix}
* & * \\
* & * \\
* & * \\
*
\end{bmatrix}
\] |

**GRAND CHALLENGE**

convex characterization in the face of structural constraints
difficult to establish relation between

\[
\begin{align*}
\{ & \text{structural constraints} \\
& \text{on } F \} \quad \text{and} \quad \{ & \text{structural constraints} \\
& \text{on } X \text{ and } Y \}
\end{align*}
\]
Optimal actuator selection

**Objective:** identify row-sparse feedback gain

\[ \text{minimize} \quad J(F) + \gamma \sum_i \| e_i^TF \|_2 \]

- variance amplification
- row-sparsity-promoting penalty function
**CHANGE OF VARIABLES:** \( Y := FX \)

- **convex dependence** of \( J \) on \( X \) and \( Y \)
- **row-sparse structure** preserved
Optimal actuator selection

 admits SDP characterization

\[
\text{minimize } J(X, Y) + \gamma \sum_{i} \|e_i^T Y\|_2
\]

\[\downarrow\]

\[\downarrow\]

variance amplification

row-sparsity-promoting penalty function

Polyak, Khlebnikov, Shcherbakov, ECC ’13
Münz, Pfister, Wolfrum, IEEE TAC ’14
Dhingra, Jovanović, Luo, CDC ’14
Sensor selection: dual problem

- **Kalman Filter**
  - minimum variance estimator

\[
\begin{align*}
\dot{x} &= Ax + L(y - \hat{y}) + Bd \\
\hat{y} &= C\hat{x} \\
y &= Cx + w
\end{align*}
\]

Objective: minimize estimation error using a few sensors

- proxy: column sparsity of Kalman gain \( L \)
Challenge: computational complexity

\[ \text{trace} \left( R Y X^{-1} Y^T \right) = \text{trace} \left( R \Theta \right) \]

worst case complexity: \( O \left( (n + m)^6 \right) \)
Customized Algorithms
Actuator selection

\[
\begin{align*}
\text{minimize} & \quad J(X, Y) + \gamma g(Y) \\
\text{subject to} & \quad AX - BY + W = 0 \\
& \quad X \succ 0
\end{align*}
\]

\[
J(X, Y) := \text{trace} \left( QX + RY^T X^{-1}Y \right)
\]

\[
g(Y) := \sum_i \|e_i^T Y\|_2
\]

\[
AX := AX + XA^T
\]

\[
BY := B_2 Y + Y^T B_2^T
\]

\[
W := B_1 B_1^T
\]
Customized algorithms

- **Alternating Direction Method of Multipliers (ADMM)**
  
  *Boyd et al., FnT in Machine Learning ’11*

- **Proximal Gradient Algorithm**
  
  *Parikh & Boyd, FnT in Optimization ’14*
Two pillars

- **AUGMENTED LAGRANGIAN**

\[
\mathcal{L}_\rho(X, Y; \Lambda) := J(X, Y) + \gamma g(Y) + \langle \Lambda, AX - BY + W \rangle + \frac{\rho}{2} \|AX - BY + W\|_F^2
\]
Two pillars

- **Augmented Lagrangian**

\[ \mathcal{L}_\rho(X, Y; \Lambda) := J(X, Y) + \gamma g(Y) + \langle \Lambda, AX - BY + W \rangle + \frac{\rho}{2} \| AX - BY + W \|_F^2 \]

- **Proximal Operator**

\[ \text{prox}_{\mu g}(V) := \arg\min_X g(X) + \frac{1}{2\mu} \| X - V \|_F^2 \]
\[
X^{k+1} := \arg\min_X \mathcal{L}_\rho(X, Y^k; \Lambda^k)
\]
\[
Y^{k+1} := \arg\min_Y \mathcal{L}_\rho(X^{k+1}, Y; \Lambda^k)
\]
\[
\Lambda^{k+1} := \Lambda^k + \rho (AX^{k+1} - BY^{k+1} + W)
\]
**$Y$-update**

\[
\text{minimize } \gamma \sum_i \|e_i^T Y\|_2 + \frac{\rho}{2} \|B Y - V\|_F^2 \underbrace{h(Y)}_{\text{group lasso}}
\]

- **Group LASSO**

\[
Y^{j+1} = \text{prox}_{\gamma \alpha^j g}(Y^j - \alpha^j \nabla h(Y^j))
\]

**soft thresholding**

\[
e_i^T Y^{j+1} = S_{\gamma \alpha^j}(e_i^T (Y^j - \alpha^j \nabla h(Y^j)))
\]

**complexity per inner iteration:** $O(nm)$
**X-update**

\[
\begin{align*}
\text{minimize} & \quad \text{trace} \left( XQ + X^{-1} Y^T RY \right) + \frac{\rho}{2} \|AX - U\|_F^2 \\
\text{subject to} & \quad X \succ 0
\end{align*}
\]

- **Can formulate as SDP**
  - worst-case complexity \(O(n^6)\)
- **Projected Newton's method**
  - use conjugate gradients to find the search direction
  - project onto \(\{X \mid X \succ 0\}\)

**Worst-case complexity:** \(O(n^5)\)

*Dhingra, Jovanović, Luo, CDC '14*
ADMM

- difficult subproblems
- slow overall convergence

**Alternative Approach**

- invertible $A$: avoid dualizing the linear constraint

\[ AX - BY + W = 0 \]
Elimination of $X$

- For invertible $A$
  
  ★ matrix $A$ doesn’t have e-values with equal positive and negative parts

$$X(Y) = A^{-1}(BY - W)$$

minimize $\gamma g(Y)$

subject to $X(Y) \succ 0$

$$J(Y) := \text{trace} \left( QX(Y) + RY^T X^{-1}(Y) Y \right)$$
Proximal gradient method

\[ Y^{k+1} := \text{prox}_{\gamma \alpha^k g} \left( Y^k - \alpha^k \nabla J(Y^k) \right) \]

soft thresholding

\[ e_i^T Y^{k+1} = S_{\gamma \alpha^k} (e_i^T (Y^k - \alpha^k \nabla J(Y^k))) \]

complexity: \( O(\max(n^3, n^2 m)) \)
• **Complexity per iteration**
  
  ★ $q$ backtracking steps: $O(q n^3)$

• **Stopping criterion**
  
  ★ terminate when relative and normalized residuals are small

*Goldstein, Studer, Baraniuk, arXiv:1411.3406*
Examples
• **OBJECTIVE**

  ★ detect aeroelastic instabilities
using half the sensors: degrades performance by $\approx 20\%$
Linearized Swift-Hohenberg equation

**PDE with spatially periodic coefficients**

\[
\partial_t \psi = -(\partial_{xx} + I)^2 \psi - c \psi + f \partial_x \psi + d + u
\]

\[
f(x) = \alpha \cos (\Omega x)
\]

where

\[
A = -(\partial_{xx} + I)^2 - cI + \alpha \cos (\Omega x) \partial_x
\]
- $n = 64; c = -0.2, \alpha = 2, \Omega = 1.25$

\[
\frac{(J - J_c)}{J_c} \frac{\gamma \text{card}(F)}{\text{card}(F_c)}
\]
Structure of optimal controller

- \( n = 64; \ c = -0.2, \ \alpha = 2, \ \Omega = 1.25 \)

**feedback gain matrix**

- \( \gamma = 0 \)
- \( \gamma = 0.675 \)

**row norms**

- 24.8% performance degradation
Comparison with ADMM

\[ \frac{\|Y^k - Y^*\|_F}{\|Y^*\|_F} \]

---

- **Proximal gradient**
- **ADMM**
Remarks

- **Convex Characterization of Sensor/Actuator Selection**
  
  Polyak, Khlebnikov, Shcherbakov, ECC ’13

- **Alternating Direction Method of Multipliers**
  
  Dhingra, Jovanović, Luo, CDC ’14

- **Proximal Gradient Algorithm**
  
  - elimination of $X$
  
  - adaptive step-size selection

- **Relation to Minimum Energy Covariance Completion Problem**
  
  - additional linear constraint on the covariance matrix $X$

  Zare, Dhingra, Jovanović, Georgiou, CDC ’17 (to appear)