Sensors measure a spatially correlated unknown field

Application Example: Smart Agriculture

Sensors are powered by a multi-antenna energy beacon and send their measurements to the sink

Main Aim:
Reconstruct the unknown field with as small average distortion as possible under a total power constraint at the energy beacon

Information Transfer

Communications to the Sink
- The signal that is received by the sink from sensor \( i \) at time slot \( t \) is
  \[
  u_i^t = r_i^t \sqrt{\frac{E_s}{2}} + w_i^t,
  \]
  where
  - \( r_i^t \) is samples of the unknown signal at sensor \( i \) at time slot \( t \)
  - \( E_s \) is power amplification factor of sensor \( i \) at time slot \( t \)
  - \( \sqrt{\frac{E_s}{2}} \) is effective channel gain for sensor \( i \) at time slot \( t \)
  - \( w_i^t \) is received observations for sensor \( i \) at time slot \( t \)
  - \( w_i^t \) is zero-mean proper white noise

Performance Criterion
Mean-square error for reconstruction of the unknown field: \( \text{MSE} = \sum_i E[(s_i - \hat{s}_i)^2] \)

Wireless Power Transfer

Energy beacon serves \( n \) sensors using a beamforming strategy \( K_{s,t} = \sum_{i=1}^{n} \gamma_i e_p \)

\[
E_i^t = \sum_{i=1}^{n} \gamma_i E_i^t,
\]

where \( E_i^t \) is energy harvested by sensor \( i \) during time slot \( t \)

Problem Statement
We jointly design optimal
- beamforming strategies \( K_{s,t} \) at the energy beacon
- power amplification factors \( p_i \) at the sensors

in order to
- minimize the MSE over the time period of \( 1 \leq t \leq n \)

\[
\begin{align*}
\min_{p, K_{s,t}} & \quad \sum_{t=1}^{n} c_t(p_i) \\
\text{s.t.} & \quad \sum_{t=1}^{n} \gamma_i E_i^t \leq \sum_{t=1}^{n} \psi_t(\gamma_i E_i^t), \quad \forall t, \forall i \\
& \quad \text{“energy neutrality constraints on sensors”} \\
& \quad \text{tr}[K_{s,t}] \leq P_b, \quad \forall t, \forall i \\
& \quad \text{“power constraint on energy beacon”}
\end{align*}
\]

Two Approaches: Reinforcement Learning vs. Optimization

**REINFORCEMENT LEARNING**
- does not rely on prior knowledge
  - no channel state information (CSI)
  - no knowledge on the form of the utility function
- does not rely on strong assumptions
  - Markovian assumption
  - feedback on the utility function and battery level information from the previous time slot is available
- does not guarantee convergence
- takes many iterations to converge
- optimize by interacting with the system

**STANDARD OPTIMIZATION**
- typically requires knowledge of system parameters (but robust solutions are also possible)
  - CSI, form of the utility function (error function) and statistics of the unknown field is known
  - may guarantee optimality if the problem is well-behaved (for instance convex)
  - our problem is convex with \( \phi_E \) and \( \phi_P \) but not with \( \phi_d \)
  - may provide convergence guarantees
  - convergence to an optimal solution is guaranteed for \( \phi_E \) and \( \phi_P \)
- no online training is required
- requires a system model

Deep Reinforcement Learning Approach

**Method:** Proximal Policy Optimization

**Reward:** negative of the MSE at each time step

**Decision variables:**
- ratio of the energy to be used to the battery level at each sensor
- energy allocated to each beamforming dictionary element at each time step at the energy beacon

Experiments

Set-up for the Experiments:
- Random Field Model: Gaussian-Schell model (GSM) with time-varying parameters
- Sensor Network: Energy beacon at \((0,1)\), sensors on the line at \(y = 0\), sink at \((0,4)\) (meters)
  - Aim: Estimate the unknown field values at \(n = 33\) positions on the line at \(y = 0\)

MSE vs. Power Budget

RL approach successfully learns to minimize the MSE without a priori knowledge of system parameters

MSE vs. number of sensors

MSE depends significantly on the number of sensors (consistent with the fact that the unknown field becomes uncorrelated periodically)

Power Allocation for Communications to the Sink

Power allocation is time varying (consistent with the time varying nature of the field correlation)

RL Convergence

With four 3.5 GHz cores and a Quadro K620 GPU,
- direct optimization and RL \(10^5\) iterations, utilizing GPU takes 15 and 62 minutes, respectively.

Acknowledgements

A. Özçelikkale acknowledges the support from Swedish Research Council under grant 2015-04011. M. Koseoglu acknowledges the support from Fulbright Program with grant number FY-2017-TR-PD-02.