ECE 598: Computational Power Systems **Winter 2025**

Lecture 11: Inverse Optimization

Lecturer: Vladimir Dvorkin *Scribe(s): Shengyang Wu & Wei Ai*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

11.1 Classic Inverse Optimization

In modern electricity markets, system operators (ISOs) solve a parameterized Optimal Power Flow (OPF) to dispatch generation and set prices. We can abstract this as:

$$\underset{\mathbf{p}}{\text{minimize}} \;\; f(\mathbf{p}, \mathbf{y}) \quad \text{subject to} \quad g(\mathbf{p}, \mathbf{y}) = \mathbf{0} \quad : \; \boldsymbol{\lambda},$$

where

- $\mathbf{p}$ is the vector of generator outputs,
- $\mathbf{y}$ encapsulates unknown parameters—e.g. wind-power forecasts, demand forecasts, marginal costs or network limits,
- $\boldsymbol{\lambda}$ are the dual multipliers, observed in practice as Locational Marginal Prices (LMPs).

Although $\mathbf{y}$ remains hidden, we may have publicly available data:

- **Real-time dispatch & LMPs:** via platforms like `gridstatus.io` and `electricitymaps.com`,
- **Weather & forecasts:** wind speed, temperature from meteorological APIs,
- **Historical market outcomes:** aggregated generation, flows, and prices from third-party analytics.

Goal: Given known input data or observed decisions, recover the latent parameters $\mathbf{y}$ (e.g. the ISO's wind forecast model or cost coefficients) that make the OPF model consistent with the recorded dispatch and prices.

11.1.1 Taxonomy of inverse optimization (IO) problems

Inverse optimization problems can be organized along three dimensions: the structure of the underlying decision model (linear, quadratic integer, non-convex), which parameters are unknown (linear, quadratic integer, non-convex), and how we measure fit between observed and predicted decisions (depending on available data, we may achieve the perfect fit or, at least, maximize a suitable measure of fitness).

11.1.2 Forward optimization (FO) problem

Original problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top C\mathbf{x} + c^\top \mathbf{x}, \\ \text{subject to} \quad & A\,\mathbf{x} \geq b. \end{aligned}$$

FO surrogate (same structure, unknown parameters):

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \Theta\mathbf{x} + \theta^\top \mathbf{x}, \\ \text{subject to} \quad & \Phi\,\mathbf{x} \geq \phi. \end{aligned}$$

- FO replicates the original structure; Θ, θ are unknown objective parameters, Φ, ϕ are unknown constraint parameters.
- To estimate the objective: fix $\Phi = A$, $\phi = b$ and recover Θ, θ.
- To estimate the constraints: fix $\Theta = C$, $\theta = c$ and recover Φ, ϕ.

11.1.3 Classical IO using bilevel programming

We consider the following scenario:

- Second-order cost coefficients in $\mathbf{C}$ and constraint matrix $\mathbf{A}$ are known. The optimal primal and dual decisions $\mathbf{x}^\star$ and $\boldsymbol{\lambda}^\star$ from the original problem are also known (from public data).
- First-order cost coefficients in $\mathbf{c}$ and the right-hand side parameter $\mathbf{b}$ are unknown

and we try to find such model with first-order coefficient $\boldsymbol{\theta}$ and right-hand side parameter $\boldsymbol{\psi}$ that can best approximate the true decisions (i.e. minimize the decision error)

$$\begin{aligned} \underset{\boldsymbol{\theta},\boldsymbol{\psi}}{\text{minimize}} \quad & \underbrace{\|\mathbf{x}(\boldsymbol{\theta},\boldsymbol{\psi}) - \mathbf{x}^\star\|_p}_{\text{primal error}} + \underbrace{\|\boldsymbol{\lambda}(\boldsymbol{\theta},\boldsymbol{\psi}) - \boldsymbol{\lambda}^\star\|_p}_{\text{dual error}} \\ \text{subject to} \quad & \underline{\boldsymbol{\theta}} \leqslant \boldsymbol{\theta} \leqslant \overline{\boldsymbol{\theta}} \\ & \underline{\boldsymbol{\psi}} \leqslant \boldsymbol{\psi} \leqslant \overline{\boldsymbol{\psi}} \\ & \mathbf{x}(\boldsymbol{\theta},\boldsymbol{\psi}) \cup \boldsymbol{\lambda}(\boldsymbol{\theta},\boldsymbol{\psi}) \in \underset{\mathbf{x},\boldsymbol{\lambda}}{\text{argmin}} \quad \mathbf{x}^\top \mathbf{C}\mathbf{x} + \boldsymbol{\theta}^\top \mathbf{x} \\ & \qquad\qquad\qquad\qquad\qquad \text{subject to} \quad \mathbf{A}\mathbf{x} \geqslant \boldsymbol{\psi} : \boldsymbol{\lambda} \end{aligned}$$

This problem can be solved by: 1) Taking the KKT conditions of the lower level problem and reformulate the whole problem as mixed-integer optimization. 2) Use stochastic gradient descent method. It is worth noting that for the first approach, we will not always get a unique solution when the cost function are linear.

11.2 Data-driven IO for static parameter estimation

The major differences between classic IO and data-driven IO lies in the number of data available. In classic IO, we only relied on a single observation $\mathbf{x}^\star$ and $\boldsymbol{\lambda}^\star$. Data-driven IO acts on the history of n observations $(\mathbf{x}_i^\star, \boldsymbol{\lambda}_i^\star), \ldots, (\mathbf{x}_n^\star, \boldsymbol{\lambda}_n^\star)$.

The whole data-driven IO formulation for static parameter (right-hand side parameter ψ in the constraints) estimation are as

$$
\begin{aligned}
\underset{\boldsymbol{\psi}}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{n}\Big(\|\mathbf{x}_i(\boldsymbol{\psi}) - \mathbf{x}_i^\star\|_p + \|\boldsymbol{\lambda}_i(\boldsymbol{\psi}) - \boldsymbol{\lambda}_i^\star\|_p\Big) && (11.1)\\
\text{subject to} \quad & \underline{\boldsymbol{\psi}} \leqslant \boldsymbol{\psi} \leqslant \overline{\boldsymbol{\psi}} && (11.2)\\
& \mathbf{x}_i(\boldsymbol{\psi}) \cup \boldsymbol{\lambda}_i(\boldsymbol{\psi}) \in \underset{\mathbf{x},\boldsymbol{\lambda}}{\text{argmin}} \quad \mathbf{x}^\top \mathbf{C}_i \mathbf{x} + \mathbf{c}_i^\top \mathbf{x} && (11.3)\\
& \qquad\qquad\qquad\qquad \text{subject to} \quad \mathbf{A}_i \mathbf{x} \geqslant \boldsymbol{\psi} : \boldsymbol{\lambda} && (11.4)
\end{aligned}
$$

The major problem here is that the ψ on the right hand side are different for every record of observation, and the computational burden would be large if we perform this optimization problem for each observation. We make the assumption that the the right hand side parameter comes from a linear prediction model $\widehat{\mathbf{b}} = \mathbf{B}\widehat{\boldsymbol{\varphi}}$, and the problem is reformulated as

$$
\begin{aligned}
\underset{\mathbf{B}}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{n}\Big(\|\mathbf{x}_i(\mathbf{B}) - \mathbf{x}_i^\star\|_p + \|\boldsymbol{\lambda}_i(\mathbf{B}) - \boldsymbol{\lambda}_i^\star\|_p\Big) && (11.5)\\
\text{subject to} \quad & \mathbf{x}_i(\mathbf{B}) \cup \boldsymbol{\lambda}_i(\mathbf{B}) \in \underset{\mathbf{x},\boldsymbol{\lambda}}{\text{argmin}} \quad \mathbf{x}^\top \mathbf{C}_i \mathbf{x} + \mathbf{c}_i^\top \mathbf{x} && (11.6)\\
& \qquad\qquad\qquad\qquad \text{subject to} \quad \mathbf{A}_i \mathbf{x} \geqslant \mathbf{B}\boldsymbol{\varphi}_i : \boldsymbol{\lambda} && (11.7)
\end{aligned}
$$

11.3 Application in Unveiling ISO's Wind Power Forecast Model

11.3.1 Problem Formulation

Here we consider an optimal power flow problem with consideration for wind power forecast $\widehat{\mathbf{w}}$

$$
\begin{aligned}
\underset{\mathbf{p}}{\text{minimize}} \quad & \mathbf{p}^\top \mathbf{C}\mathbf{p} + \mathbf{c}^\top \mathbf{p} && \textit{generator dispatch cost}\\
\text{subject to} \quad & \underline{\mathbf{p}} \leqslant \mathbf{p} \leqslant \overline{\mathbf{p}} && \textit{generation limits}\\
& \mathbf{1}^\top(\mathbf{p} + \widehat{\mathbf{w}} - \mathbf{d}) = 0 \quad : \boldsymbol{\lambda}_b && \textit{power balance condition}\\
& |\mathbf{F}(\mathbf{p} + \widehat{\mathbf{w}} - \mathbf{d})| \leqslant \overline{\mathbf{f}} \quad : \boldsymbol{\lambda}_{\overline{\mathbf{f}}}, \boldsymbol{\lambda}_{\underline{\mathbf{f}}} && \textit{power flow limits}
\end{aligned}
$$

Our goal is to get the parameters for ISO wind power forecast model with available public dataset, including public weather data $\boldsymbol{\varphi}$ and Locational Marginal Prices (LMP) $\boldsymbol{\pi} = \boldsymbol{\lambda}_b \cdot \mathbf{1} - \mathbf{F}^\top(\boldsymbol{\lambda}_{\overline{\mathbf{f}}} - \boldsymbol{\lambda}_{\underline{\mathbf{f}}})$

11.3.2 IO Problem Solving

We first reformulate the problem in the FO form:

$$\begin{aligned} \underset{\mathbf{p}}{\text{minimize}} \quad & \frac{1}{2}\mathbf{p}^\top \mathbf{C}\mathbf{p} + \mathbf{c}^\top \mathbf{p} \\ \text{subject to} \quad & \mathbf{A}\mathbf{p} \geqslant \mathbf{b}(\widehat{\mathbf{w}}) \quad : \boldsymbol{\lambda} \end{aligned}$$

and then take its dual

$$\begin{aligned} \underset{\boldsymbol{\lambda}}{\text{maximize}} \quad & \mathbf{q}(\widehat{\mathbf{w}})^\top \boldsymbol{\lambda} - \boldsymbol{\lambda}^\top \mathbf{Q}\boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geqslant \mathbf{0} \end{aligned}$$

where $\mathbf{q}(\widehat{\mathbf{w}}) = \mathbf{A}\mathbf{C}^{-1}\mathbf{c} + \mathbf{b}\widehat{\mathbf{w}})$ and $\mathbf{Q} = \mathbf{A}\mathbf{C}^{-1}\mathbf{A}$

By solving this optimization problem which is parameterized by $\widehat{\mathbf{w}}$, we can get the estimated LMP $\boldsymbol{\pi}(\widehat{\mathbf{w}})$. By minimizing the difference between the estimated LMP $\boldsymbol{\pi}(\widehat{\mathbf{w}})$ and the true LMP $\boldsymbol{\pi}^*$ from historic observations using the data driven IO model, the wind forecast model parameter $\mathbf{B}$ can be estimated.