Optimal Design:
\[
\max_u \ln \det(XUX^T) \text{ s.t. } e^Tu = 1, \ u \geq 0.
\]
i.e. “minimize the (determinant of the) variance of the estimator \( \hat{\theta} \) by choosing each \( x_i \) with probability \( u_i \).”

An alternative formulation is
\[
\max_{u,A} \ln \det A \\
A - XUX^T = 0 \\
e^Tu = 1, \ u \geq 0, \\
A \text{ symmetric and positive semidefinite.}
\]

We have an inner product on \( m \times n \) matrices:
\[
A \cdot B = \langle A,B \rangle := \text{trace} (A^TB) = \sum_i \sum_j a_{ij}b_{ij}.
\]

Now look at the “derivative” of \( \phi(A) := \ln \det A \), defined on \( n \times n \) matrices with positive determinant. Let
\[
f(\lambda) : = \ln \det(A + \lambda e_ie_j^T) \\
= \ln(\det A \det(1 + \lambda e_i^T A^{-1} e_i)) \\
= \ln \det A + \ln(1 + \lambda e_i^T A^{-1} e_i).
\]

So \( f'(0) = e_j^T A^{-1} e_i = \text{trace} (A^{-1} \cdot (e_i e_j^T)) = A^{-T} \cdot (e_i e_j^T) \). In general, the directional derivative of \( \phi \) at \( A \) in the direction \( D \) is
\[
\phi'(A; D) = \lim_{\lambda \to 0} \frac{\phi(A + \lambda D) - \phi(A)}{\lambda} = A^{-T} \cdot D.
\]

In particular, if \( A \) and \( D \) are symmetric, we get \( \phi'(A; D) = A^{-1} \cdot D \) and can write \( \nabla \phi(A) = A^{-1} \).

[Note: If \( \psi(A) = \phi'(A; D) \) and \( E \) is also symmetric, then
\[
\psi'(A; E) = -(A^{-1} EA^{-1}) \cdot D = -(A^{-1} DA^{-1}) \cdot E.
\]

This can be used to show that \( -\ln \det \) is convex on the space of symmetric positive definite matrices.]

Now consider the Lagrangian dual of the reformulated optimal design problem, which is equivalent to
\[
\max_{u \geq 0, A \text{ symm., p.d.}} \min_{H \text{ symm., } \lambda \in \mathbb{R}} \{ \ln \det A - H \cdot (A - XUX^T) - \lambda(e^Tu - 1) \}.
\]
The dual is
\[
\min_{H \text{ symm.}, \lambda \in \mathbb{R}} \max_{u \geq 0, A \text{ symm., p.d.}} \left\{ \ln \det A - H \cdot A + [H \cdot XUX^T - \lambda e^T u] + \lambda \right\}.
\]

Note that
\[
H \cdot (XUX^T) = H \cdot \left( \sum_{i=1}^{m} u_i x_i x_i^T \right) = \sum_{i=1}^{m} u_i H \cdot x_i x_i^T = \sum_{i=1}^{m} u_i (x_i^T H x_i).
\]

So the maximum over nonnegative \(u\)'s is \(+\infty\) unless \(x_i^T H x_i \leq \lambda\) for all \(i\), in which case it is zero; so the dual can be written
\[
\min_{H \text{ symm.}, \lambda \in \mathbb{R}, x_i^T H x_i \leq \lambda, i=1, \cdots, m} \left\{ \lambda + \max_{A \text{ symm., p.d.}} (\ln \det A - H \cdot A) \right\}.
\]

Now using our discussion of the derivative of \(\ln \det\) we find that the maximum is achieved by \(A = H^{-1}\) (and it is not hard to see that the maximum is \(\infty\) if \(H\) is not positive definite). So the dual becomes
\[
\min_{H \text{ symm., p.d.}, \lambda > 0, x_i^T H x_i \leq \lambda, i=1, \cdots, m} \left\{ \lambda - \ln \det H - n \right\}.
\]

Now rewrite this in terms of
\[
M = \frac{n}{\lambda} H \Rightarrow H = \frac{\lambda}{n} M.
\]

Then we get the following equivalent problem
\[
\min_{M \text{ symm., p.d.}, \lambda > 0, x_i^T M x_i \leq n, i=1, \cdots, m} \left\{ \lambda - n \ln \lambda + n \ln n - \ln \det M - n \right\}.
\]

Since the above function is convex with respect to \(\lambda\), then we can optimize over \(\lambda\) to find \(\lambda = n\) and so get the following equivalent problem:
\[
\min_{M \text{ symm., p.d.}} - \ln \det M \quad \text{s.t.} \quad x_i^T M x_i \leq n, i = 1, \cdots, m.
\]

So the dual problem is that of finding the minimum-volume ellipsoid centered at 0 containing \(X = \{x_1, \ldots, x_m\}\).

In fact, this dual lends some “robustness” to the notion of D-optimality. \(\frac{\sigma^2}{N}(XUX^T)^{-1}\) is the variance-covariance matrix of \(\hat{\theta}\). The variance of the prediction \(\hat{\theta}^T x_i\) at a design point \(x_i\) is \(\frac{\sigma^2}{N} x_i^T (XUX^T)^{-1} x_i\). We might want to minimize
\[
\max_{i} x_i^T (XUX^T)^{-1} x_i.
\]

This is minimized by the same \(u\) as solves the D-optimal design problem.
(i) The optimal value is at least \( n \), since

\[
\max_i x_i^T (UXU^T)^{-1} x_i \geq \sum_i u_i x_i^T (UXU^T)^{-1} x_i
\]

\[
= \text{trace} \left( \sum_i u_i x_i^T \cdot (UXU^T)^{-1} \right)
\]

\[
= \text{trace} \left( (UXU^T) \cdot (UXU^T)^{-1} \right) = n.
\]

(ii) It is not hard to show that if \( u \) is D-optimal, \( H = (UXU^T)^{-1} \) is optimal in the dual, so feasible in the dual, so \( x_i^T H x_i \leq n \) for all \( i \).

In this course, we looked at various extensions of LP. The following is a brief review of this course:

1. LCPs: we still have pivoting algorithms for these, but no objective function to force convergence; instead we have a combinatorial convergence proof. No explicit convexity, but convexity related to the monotonicity of \( M \).

2. Complexity of pivoting algorithms: neighborly polytopes, diameters of polytopes (false Hirsch conjecture), exponential examples, polynomial expected behavior of a special pivoting algorithm, smoothed complexity.

3. Complexity of NLP: impossibility of approximating the minimum of a nonconvex function or the minimizer of a convex function, complexity of approximating the minimum of a convex functions, either nonsmooth functions (MCG, ellipsoid method, method of inscribed ellipsoids and the subgradient method), or smooth functions (gradient method).

4. Equivalence of separation and optimization: using the ellipsoid method and polarity; motivates cutting-plane methods.

5. Duality: conic programming duality for regression, data classification and robust LP; Lagrangian duality for the D-optimal design problem.