M677_W14

November 20, 2023

1 Week 14: Semidefinite Programming

Last updated: 20 November 2023

We first recall that a matrix $X \in \mathbb{R}^{d\times d}$ is called positive semidefinite if it is symmetric and if $\langle Xv, v\rangle \geq 0$ for all $v \in \mathbb{R}^d$. Equivalently, by the spectral theorem, a matrix is positive semidefinite if it is symmetric and if all its eigenvalues are nonnegative. To abbreviate that $X \in \mathbb{R}^{d\times d}$ is positive semidefinite, we write $X \succeq 0$. The notation $X \succeq X'$ or $X' \preceq X$ means that $X' - X \succeq 0$.

A semidefinite program in standard form is a convex optimization program that reads

$$\underset{X\in\mathbb{R}^{d\times d}}{\text{minimize}}\ \mathrm{tr}(C^\top X) \qquad \text{subject to } \mathrm{tr}(A_i^\top X) = b_i, i \in \{1,\dots,m\},\ X \succeq 0,$$

for some matrices $A_1, \dots, A_m, C \in \mathbb{R}^{d\times d}$—which can be assumed to be symmetric. Note the similarity with the standard form of linear programs, in particular in view of the interpretation of e.g. $\mathrm{tr}(C^\top X)$ as the the Frobenius inner product $\langle C, X\rangle_F$. Semidefinite programs encapsulate linear programs, since a linear program in standard form can be written as

$$\underset{X\in\mathbb{R}^{d\times d}}{\text{minimize}}\ \langle \mathrm{diag}[c], X\rangle \qquad \text{subject to } \mathrm{tr}(\mathrm{diag}[a_i]X) = b_i,\ \mathrm{tr}(E_{k,\ell}X) = 0,\ X \succeq 0.$$

In general, a semidefinite program is any optimization program mixing vector or matrix variables and involving linear inequality constraints in the regular or semidefinite sense. We take for granted from now on that semidefinite programs can be solved effciently.

Example: The optimization program (which could be solved explicitely)

$$\underset{x\in\mathbb{R}^{d}}{\text{minimize}}\ \|Mx\|_2 \qquad \text{subject to } Ax = b$$

can be reformulated as the semidefinite program

$$\underset{x\in\mathbb{R}^{d},\, c\in\mathbb{R}}{\text{minimize}}\ c \qquad \text{subject to } Ax = b \ \text{ and to } \left[\begin{array}{c|c} cI_d & Mx \\ \hline (Mx)^\top & c \end{array}\right] \succeq 0$$

This is done by introducing the slack variable c such that $\|Mx\|_2 \leq c$ and remarking that the latter is equivalent to $\left[\begin{array}{c|c} cI_d & Mx \\ \hline (Mx)^\top & c \end{array}\right] \succeq 0$.

Example: The operator norm $\|M\|_{\mathrm{op}}$ of a matrix $M \in \mathbb{R}^{d\times d}$ is the solution to the semidefinite program

$$\underset{c \in \mathbb{R}}{\text{minimize}}\ c \qquad \text{subject to } \left[\begin{array}{c|c} 0 & M \\ \hline M^\top & 0 \end{array}\right] \preceq cI_{2d}.$$

There is a duality theory for semidefinite programming (and for conic programing in general). Here is a representative result.

Theorem: The primal and dual semidefinite programs

$$\begin{aligned} \underset{X \in \mathbb{R}^{d \times s}}{\text{minimize}}\ \mathrm{tr}(C^\top X) \quad & \text{subject to } \mathrm{tr}(A_i^\top X) = b_i, i \in \{1, \ldots, m\},\ X \succeq 0, \\ \underset{\nu \in \mathbb{R}^m}{\text{maximize}}\ \langle -b, \nu \rangle \quad & \text{subject to } \nu_1 A_1 + \cdots + \nu_m A_m + C \succeq 0, \end{aligned}$$

have equal optimal values, provided the primal program or the dual program is strongly feasible.

Example: The nuclear norm $\|M\|_* := \sum_{i=1}^d \sigma_i(M)$ of a matrix $M \in \mathbb{R}^{d \times d}$ is the solution to the semidefinite program

$$\underset{P,Q \in \mathbb{R}^{d \times d}}{\text{minimize}}\ \frac{1}{2}\Big(\mathrm{tr}(P) + \mathrm{tr}(Q)\Big) \qquad \text{subject to } \left[\begin{array}{c|c} P & M \\ \hline M^\top & Q \end{array}\right] \succeq 0.$$

Proof: See details in class. The first step is the observation that

$$\|M\|_* = \max\{\mathrm{tr}(AM) : A \in \mathbb{R}^{d \times d} \text{ satisfies } \|A\|_{\mathrm{op}} \leq 1\}.$$

The second step consists in using semidefinite duality.

As an application, a low-rank matrix $M \in \mathbb{R}^{d \times d}$ can be recovered by nuclear-norm minimization from the knowledge of only a few of its entries. See the MATLAB illustration. The theoretical justification belongs to the field of Compressive Sensing (which will be discussed next week).