8.3 [Optional] Ivanov implies Tikhonov for Ridge Regression.

To show that Ivanov implies Tikhonov for the ridge regression problem (square loss with ℓ_2 regularization), we need to demonstrate strong duality and that the dual optimum is attained. Both of these things are implied by Slater's constraint qualifications.

1. [Optional] Show that the Ivanov form of ridge regression

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} \left(y_i - w^T x_i\right)^2 \\ \text{subject to} & w^T w \le r. \end{array}$$

   is a convex optimization problem with a strictly feasible point, so long as $r > 0$. (Thus implying the Ivanov and Tikhonov forms of ridge regression are equivalent when $r > 0$.)

   Solution: The Ivanov form of ridge regression is

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} \left(y_i - w^T x_i\right)^2 \\ \text{subject to} & w^T w \le r. \end{array}$$

   First we show this is a convex optimization problem: The objective is convex in w since we have a nonnegative combination of convex functions $\left(y_i - w^T x_i\right)^2$. Each of these expressions is convex since the square is a convex function, and we're applying it to an affine transformation of w. The constraint function is also a convex function of w. Slater's constraint qualification is satisfied since we can take $w = 0$, so long as $r > 0$. Thus we have strong duality.

A Positive Semidefinite Matrices

In statistics and machine learning, we use positive semidefinite matrices a lot. Let's recall some definitions from linear algebra that will be useful here:

Definition. A set of vectors $\{x_1, \ldots, x_n\}$ is **orthonormal** if $\langle x_i, x_i \rangle = 1$ for any $i \in \{1, \ldots, n\}$ (i.e. x_i has unit norm), and for any $i, j \in \{1, \ldots, n\}$ with $i \neq j$ we have $\langle x_i, x_j \rangle = 0$ (i.e. x_i and x_j are orthogonal).

Note that if the vectors are column vectors in a Euclidean space, we can write this as $x_i^T x_j = 1(i \neq j)$ for all $i, j \in \{1, \ldots, n\}$.

Definition. A matrix is **orthogonal** if it is a square matrix with orthonormal columns.

It follows from the definition that if a matrix $M \in \mathbf{R}^{n \times n}$ is orthogonal, then $M^T M = I$, where I is the $n \times n$ identity matrix. Thus $M^T = M^{-1}$, and so $M M^T = I$ as well.

Definition. A matrix M is **symmetric** if $M = M^T$.

Definition. For a square matrix M, if $Mv = \lambda v$ for some column vector v and scalar λ, then v is called an **eigenvector** of M and λ is the corresponding **eigenvalue**.

Theorem (Spectral Theorem). *A real, symmetric matrix $M \in \mathbf{R}^{n\times n}$ can be diagonalized as $M = Q\Sigma Q^T$, where $Q \in \mathbf{R}^{n\times n}$ is an orthogonal matrix whose columns are a set of orthonormal eigenvectors of M, and Σ is a diagonal matrix of the corresponding eigenvalues.*

Definition. A real, symmetric matrix $M \in \mathbf{R}^{n\times n}$ is **positive semidefinite (psd)** if for any $x \in \mathbf{R}^n$,

$$x^T M x \geq 0.$$

Note that unless otherwise specified, when a matrix is described as positive semidefinite, we are implicitly assuming it is real and symmetric (or complex and Hermitian in certain contexts, though not here).

As an exercise in matrix multiplication, note that for any matrix A with columns $a_1, \ldots, a_d$, that is

$$A = \begin{pmatrix} | & & | \\ a_1 & \cdots & a_d \\ | & & | \end{pmatrix} \in \mathbf{R}^{n\times d},$$

we have

$$A^T M A = \begin{pmatrix} a_1^T M a_1 & a_1^T M a_2 & \cdots & a_1^T M a_d \\ a_2^T M a_1 & a_2^T M a_2 & \cdots & a_2^T M a_d \\ \vdots & \vdots & \cdots & \vdots \\ a_d^T M a_1 & a_d^T M a_2 & \cdots & a_d^T M a_d \end{pmatrix}.$$

So M is psd if and only if for any $A \in \mathbf{R}^{n\times d}$, we have $\operatorname{diag}(A^T M A) = \left(a_1^T M a_1, \ldots, a_d^T M a_d\right)^T \succeq 0$, where $\succeq$ is elementwise inequality, and 0 is a $d \times 1$ column vector of 0's .

1. Use the definition of a psd matrix and the spectral theorem to show that all eigenvalues of a positive semidefinite matrix M are non-negative. [Hint: By Spectral theorem, $\Sigma = Q^T M Q$ for some Q. What if you take $A = Q$ in the "exercise in matrix multiplication" described above?]
   Solution: Let $M \in \mathbf{R}^{n\times n}$ be a psd matrix. By the spectral theorem, we can write $M = Q\Sigma Q^T$. Then by the argument above,

   $$\operatorname{diag}(Q^T M Q) = \operatorname{diag}(\Sigma) \succeq 0.$$

2. In this problem, we show that a psd matrix is a matrix version of a non-negative scalar, in that they both have a "square root". Show that a symmetric matrix M can be expressed as $M = BB^T$ for some matrix B, if and only if M is psd. [Hint: To show $M = BB^T$ implies M is psd, use the fact that for any vector v, $v^T v \geq 0$. To show that M psd implies $M = BB^T$ for some B, use the Spectral Theorem.]
   Solution:

   If we can write $M = BB^T$, then M is symmetric and $x^T M x = x^T B B^T x = (B^T x)^T (B^T x) = ||B^T x||^2 \geq 0$, for any x. Thus A is symmetric positive semidefinite. Conversely, if M is psd, then by the Spectral Theorem there exists Q orthogonal and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_d)$ such that $M = Q\Sigma Q^T$. By the previous problem $\sigma_i \geq 0$, so we can define $\Sigma^{1/2} := \operatorname{diag}(\sigma_1^{1/2}, \ldots, \sigma_d^{1/2})$ and write

   $$\begin{aligned} M &= Q\Sigma^{1/2}\Sigma^{1/2}Q^T \\ &= BB^T \end{aligned}$$

where $B := Q\Sigma^{1/2}$. Note that we can also write

$$M = Q\Sigma^{1/2}Q^T Q\Sigma^{1/2}Q^T = C^2,$$

where $C = Q\Sigma^{1/2}Q^T$. So M has a "square root", just like non-negative real numbers.

B Positive Definite Matrices

Definition. A real, symmetric matrix $M \in \mathbf{R}^{n\times n}$ is **positive definite** (spd) if for any $x \in \mathbf{R}^n$ with $x \neq 0$,

$$x^T M x > 0.$$

1. Show that all eigenvalues of a symmetric positive definite matrix are positive. [Hint: You can use the same method as you used for psd matrices above.]

2. Let M be a symmetric positive definite matrix. By the spectral theorem, $M = Q\Sigma Q^T$, where Σ is a diagonal matrix of the eigenvalues of M. By the previous problem, all diagonal entries of Σ are positive. If $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$, then $\Sigma^{-1} = \operatorname{diag}\left(\sigma_1^{-1}, \ldots, \sigma_n^{-1}\right)$. Show that the matrix $Q\Sigma^{-1}Q^T$ is the inverse of M.

3. Since positive semidefinite matrices may have eigenvalues that are zero, we see by the previous problem that not all psd matrices are invertible. Show that if M is a psd matrix and I is the identity matrix, then $M + \lambda I$ is symmetric positive definite for any $\lambda > 0$, and give an expression for the inverse of $M + \lambda I$.

4. Let M and N be symmetric matrices, with M positive semidefinite and N positive definite. Use the definitions of psd and spd to show that $M + N$ is symmetric positive definite. Thus $M + N$ is invertible. (Hint: For any $x \neq 0$, show that $x^T(M + N)x > 0$. Also note that $x^T(M + N)x = x^T Mx + x^T Nx$.)