考虑通过参数范数正则化的代价函数：

$$\tilde{J}(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) = J(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) + \alpha\Omega(\boldsymbol{\theta}). \tag{7.25}$$

回顾第4.4节我们可以构造一个广义 Lagrange 函数来最小化受约束的函数，即在原始目标函数加上一系列惩罚项。每个惩罚是一个系数之间的乘积，称为**Karush-Kuhn-Tucker** (Karush-Kuhn-Tucker) 乘子，以及一个表示约束是否满足的函数。如果我们想约束 $\Omega(\boldsymbol{\theta})$ 小于某个常数 k，我们可以构建广义 Lagrange 函数

$$\mathcal{L}(\boldsymbol{\theta}, \alpha; \boldsymbol{X}, \boldsymbol{y}) = J(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) + \alpha(\Omega(\boldsymbol{\theta}) - k). \tag{7.26}$$

这个约束问题的解由下式给出

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \max_{\alpha, \alpha \geq 0} \mathcal{L}(\boldsymbol{\theta}, \alpha). \tag{7.27}$$

如第4.4节中描述，解决这个问题需要同时改变 $\boldsymbol{\theta}$ 和 α。第4.5节给出了一个带 L^2 约束的线性回归实例。许多不同的优化过程是可能的，有些可能会利用梯度下降而其他可能使用梯度为 0 的解析解，但在所有程序中 α 在 $\Omega(\boldsymbol{\theta}) > k$ 时必须增加，在 $\Omega(\boldsymbol{\theta}) < k$ 时必须减小。所有正的 α 鼓励 $\Omega(\boldsymbol{\theta})$ 收缩。最佳值 α^* 也将鼓励 $\Omega(\boldsymbol{\theta})$ 收缩，但不会如 $\Omega(\boldsymbol{\theta})$ 小于 k 时那么强烈。

为了洞察约束的影响，我们可以固定 α^*，把这个问题看成只跟 $\boldsymbol{\theta}$ 有关的函数：

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \alpha^*) = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) + \alpha^*\Omega(\boldsymbol{\theta}). \tag{7.28}$$

这和最小化 $\tilde{J}$ 的正则化训练问题是完全一样的。因此，我们可以把参数范数惩罚看作对权重强加的约束。如果 Ω 是 L^2 范数，那么权重就是被约束在一个 L^2 球中。如果 Ω 是 L^1 范数，那么权重就是被约束在一个 L^1 范数限制的区域中。通常我们不知道权重衰减系数 α^* 约束的区域大小，因为 α^* 的值不直接告诉我们 k 的值。原则上我们可以解得 k，但 k 和 α^* 之间的关系取决于 J 的形式。虽然我们不知道约束区域的确切大小，但我们可以通过增加或者减小 α 来大致扩大或收缩约束区域。较大的 α，将导致一个较小的约束区域。较小的 α，将导致一个较大的约束区域。