

在某些情况下，将一个优化问题分解成几个部分，可以更快地解决原问题。^{what?}如果我们最小化 $f(\mathbf{x})$ 相对于某个单一变量 x_i ，然后相对于另一个变量 x_j ，等等，反复循环所有的变量，我们会保证到达（局部）极小值。这种做法被称为坐标下降 (coordinate descent)，因为我们一次优化一个坐标。更一般地，块坐标下降 (block coordinate descent) 是指对于某个子集的变量同时最小化。术语“坐标下降”通常既指块坐标下降，也指严格的单个坐标下降。

^{when?} 当优化问题中的不同变量能够清楚地分成相对独立的组，或是当优化一组变量明显比优化所有变量效率更高时，坐标下降最有意义。例如，考虑代价函数

$$J(\mathbf{H}, \mathbf{W}) = \sum_{i,j} |H_{i,j}| + \sum_{i,j} (\mathbf{X} - \mathbf{W}^\top \mathbf{H})_{i,j}^2. \quad (8.38)$$



