CS 395T: Homework 5

Brady Zhou
brady.zhou@utexas.edu

December 10, 2017

1 Linear Programming (Simplex Method)

We are trying to optimize the following

$$\min_x c^\intercal x \quad \text{subject to } Ax = b \tag{1}$$

The first method we use is the Simplex method on synthetic data generated by $A = rand(100, 200)$, $b = rand(100)$, and $c = rand(200)$. The objective was modified in a way such that x would have a known basic feasible solution.

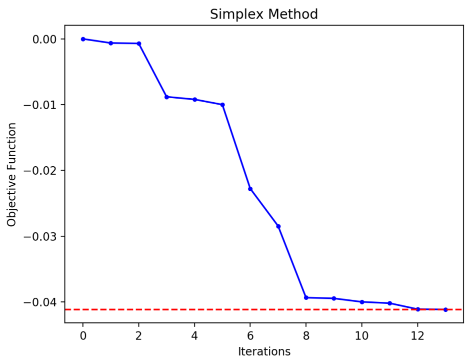

Figure 1: Simplex method.

The dotted red line is the solution from a library SciPy.

2 Linear Programming (Interior Point Method)

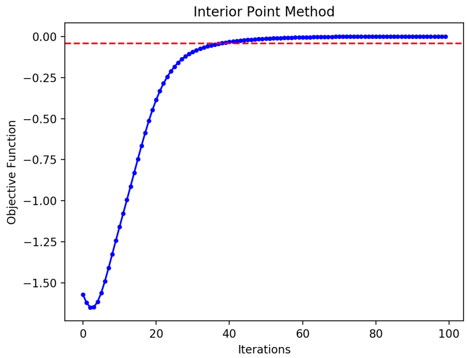

Figure 2: Interior point method.

The graph displayed is the first 100 of 1000 iterations. The interior point method initial solution does not start off as a basic feasible solution but eventually converges to a feasible point, and the same objective value as the Simplex method.

3 Matrix Approximation

We are trying to approximate a symmetric matrix $A \in \mathbb{R}^{n \times n}$, where $A = U\Sigma U^\intercal$ with a matrix $X = UBU^\intercal \succeq 0$.

$$\begin{aligned}\min_{B} ||UBU^\intercal - U\Sigma U^\intercal||_F^2 &= \min_{B\succeq 0} ||B - \Sigma||_F^2 \\ &= \min_{B\succeq 0} Tr(\Sigma^\intercal \Sigma - 2\Sigma^\intercal B + B^\intercal B) \\ &\cong \min_{B\succeq 0} -2Tr(\Sigma^\intercal B) + Tr(B^\intercal B)\end{aligned}$$

Taking the derivative with respect to b_{ij}, where $i \neq j$, we have the following

$$\begin{aligned}\frac{\partial}{\partial b_{ij}} - 2Tr(\Sigma^\intercal B) + Tr(B^\intercal B) &= 2b_{ij} \\ &= 0\end{aligned}$$

And taking the derivative with respect to b_{ij}, where $i = j$, we have

$$\begin{aligned}\frac{\partial}{\partial b_{ii}} - 2Tr(\Sigma^\intercal B) + Tr(B^\intercal B) &= -2\sigma_i + 2b_{ii} \\ &= 0\end{aligned}$$

These two results give us that $b_{ij} = 0$ if $i \neq j$, and $b_{ii} = \sigma_i$, however, to satisfy the constraint that $B \succeq 0$, we must enforce the diagonals are all non-negative, so we clip $b_{ii} \geq 0$. This gives us our result that $X = Umax(\Sigma, 0)U^\intercal$.

4 Matrix Approximation (cont.)

Given a matrix $A \in \mathbb{R}^{m\times n}$, we are looking for an r rank approximation. In this proof we assume that **all σ_i are unique**. First, we will solve the case where $r = 1$. Let $\alpha \in \mathbb{R}^m$, $\beta \in \mathbb{R}^n$.

$$\begin{aligned}
\min_{\alpha,\beta} ||U\Sigma V^\intercal - U\alpha\beta^\intercal V^\intercal||_F^2 &= \min_{\alpha,\beta} ||\Sigma - \alpha\beta||_F^2 \\
&= \min_{\alpha,\beta} Tr(\Sigma^\intercal\Sigma - 2\Sigma\alpha\beta^\intercal + \beta\alpha^\intercal\alpha\beta^\intercal) \\
&\cong \min_{\alpha,\beta} -2Tr(\Sigma\alpha\beta^\intercal) + Tr(\beta\alpha^\intercal\alpha\beta^\intercal)) \\
&= \min_{\alpha,\beta} -2\beta^\intercal\Sigma^\intercal\alpha + ||\alpha||^2||\beta||^2
\end{aligned}$$

Taking derivatives with respect to α and β,

$$\begin{aligned}
\frac{\partial}{\partial\alpha} - 2\beta^\intercal\Sigma^\intercal\alpha + ||\alpha||^2||\beta||^2 &= -2\Sigma\beta + 2||\beta||^2\alpha \\
0 &= -2\Sigma\beta + 2||\beta||^2\alpha \\
\alpha &= \frac{1}{||\beta||^2}\Sigma\beta
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\beta} - 2\beta^\intercal\Sigma^\intercal\alpha + ||\alpha||^2||\beta||^2 &= -2\Sigma^\intercal\alpha + 2||\alpha||^2\beta \\
0 &= -2\Sigma^\intercal\alpha + 2||\alpha||^2\beta \\
\beta &= \frac{1}{||\alpha||^2}\Sigma^\intercal\alpha
\end{aligned}$$

So, if $\beta \neq 0$, we have

$$\begin{aligned}
\alpha^\intercal\alpha &= \frac{1}{||\beta||^4}\beta^\intercal\Sigma^\intercal\Sigma\beta \\
\beta &= \frac{||\beta||^2}{\beta^\intercal\Sigma^\intercal\Sigma\beta}\Sigma^\intercal\Sigma\beta \\
\beta\beta^\intercal\Sigma^\intercal\Sigma\beta &= ||\beta||^2\Sigma^\intercal\Sigma\beta \\
(I - \hat{\beta}\hat{\beta}^\intercal)\Sigma^\intercal\Sigma\beta &= 0
\end{aligned}$$

This implies that, for some scalar c, $\Sigma^\intercal\Sigma\beta = c\beta$. If $\sigma_i \neq \sigma_j$, for all $i \neq j$, then β is one-sparse. Similarly, this holds for α. Now we seek to minimize the original objective function, and it is clear that $\alpha\beta^\intercal = \Sigma_1$ is optimal since we know $\alpha\beta^\intercal$ has at most 1 non-zero entry.

Back to the original question, now we seek estimate a rank r matrix, that is, $\alpha \in \mathbb{R}^{m\times r}$, and $\beta \in \mathbb{R}^{n\times r}$, where $\alpha_i \perp \alpha_j$, and $\beta_i \perp \beta_j$, and $||\alpha_i|| = ||\beta_i|| = 1$.

$$\begin{aligned}
\min_{\alpha,\beta} ||U\Sigma V^\intercal - U\alpha\beta^\intercal V^\intercal||_F^2 &= \min_{\alpha,\beta} ||\Sigma - \alpha\beta^\intercal||_F^2 \\
&= \min_{\alpha,\beta} tr(\Sigma^\intercal\Sigma - 2\Sigma^\intercal\alpha\beta^\intercal + \beta\alpha^\intercal\alpha\beta^\intercal) \\
&\cong \min_{\alpha,\beta} -2Tr(\Sigma^\intercal\alpha\beta^\intercal) + Tr(\beta\alpha^\intercal\alpha\beta^\intercal) \\
&= \min_{\alpha,\beta} -2\sum_{i=1}^{r}\beta_i^\intercal\Sigma\alpha_i + \sum_{i=1}^{r}\alpha_i^\intercal\alpha_i\beta_i^\intercal\beta_i \\
&= \min_{\alpha,\beta} \sum_{i=1}^{r} -2\beta_i^\intercal\Sigma^\intercal\alpha_i + \alpha_i^\intercal\alpha_i\beta_i^\intercal\beta_i
\end{aligned}$$

So we have reduced this problem into the rank 1 case, given all the σ_i are unique. We have to keep in mind that this problem is constrained to have each vector be orthogonal, but choosing the α_i, β_i to be one-sparse in different dimensions easily satisfies this constraint.