Serhat Selcuk Bucak
09/2007

NONLINEAR OPTIMIZATION PROBLEM (NLP)

Aim: To minimize $f(x)$ over $x \in R^n$ subject to $h(x) = 0$ (equality constraint) and $g(x) \le 0$ (inequality constraint).

For non-negative matrix factorization, the objective function is $f(x) = tr\left[\mathbf{V}^T\mathbf{V} - 2\mathbf{V}^T\mathbf{WH} + \mathbf{H}^T\mathbf{W}^T\mathbf{WH}\right]$. The equality constraint does not exist while the inequality constraints are $-\mathbf{W} \le 0$ and $-\mathbf{H} \le 0$ (component-wise). This corresponds to "single inequality constraint" situation.

Local Minimum: If $x^* \in F$, there is a neighborhood $\vartheta(x^*) \subset R^n$ s.t. $f(x) > f(x^*)$ for all $x \in \vartheta(x^*) \cap F$, where $F \triangleq \{x \in R^n \mid h(x) = 0, g(x) \le 0\}$.

Isolated Local Minimum: If x^* is the only local minimum in $\vartheta(x^*) \cap F$, it is called as the isolated local minimum.

$$\nabla f(x^*) = -\lambda^* \nabla h(x^*) \tag{1}$$

Eq. (1) is a necessary condition for optimality in the general case.

Single Equality Condition

- Taylor expression of $h(x+d)$ is given in Eq.(2) for $d \in R^n$:

$$h(x+d) \approx h(x) + \nabla h(x)^T d \tag{2}$$

Since $h(x) = 0$ due to the equality constraint, we have $h(x+d) \approx \nabla h(x)^T d$. The equality constraint should be satisfied for $h(x+d)$ as well. Thus we get:

$$\nabla h(x)^T d = 0 \tag{3}$$

- On the other hand, we want $f(x) < f(x+d)$ (minimizing the cost function) for $d \in R^n$.

$$0 > f(x+d) - f(x) \approx \nabla f(x)^T d \text{ (Taylor series)} \tag{4}$$

$$0 > \nabla f(x)^T d \tag{5}$$

- If x is a local minimum, a "direction d" which satisfies both Eq. (5) and Eq. (3) cannot exist. For all d, Eq. (3) should be satisfied because of the equality constraint. However Eq. (5) is a contradiction to the definition of a local minimum. This is because if x^* is a local minimum, then the expression $f(x^*) < f(x^* + d)$ should not be valid. Therefore in order to find the local minimum, we want a solution for the direction d which should make the simultaneous solution to Eq. (5) and Eq. (3) indefinite. So the direction d which contradicts the optimization procedure will give us the solution for the local minimum. The condition of $\nabla f(x)$ is being parallel to $\nabla h(x)$ is the only situation that satisfies the above requirement. From Eq. (3), the orthogonality between $\nabla h(x)$ and *d* is straightforward. If $\nabla f(x)$ and $\nabla h(x)$ are parallel, then Eq.(5) cannot be solved, which also means that the condition to have the local minimum is satisfied.

In addition to introducing the Lagrangian function in Eq.(6), by using Eq. (1) and $\nabla_x L(x,\lambda) = \nabla f(x) + \lambda \nabla h(x)$ we get Eq. (7).

$$L(x,\lambda) = f(x) + \lambda h(x) \tag{6}$$

$$\nabla_x L(x^*,\lambda^*) = 0 \tag{7}$$

Single Inquality Condition

E.g. $\begin{array}{l} f(x) = x_1 + x_2, \quad x = (x_1, x_2) \in R^2 \\ g(x) \triangleq x_1^2 + x_2^2 - 2 \le 0 \qquad (\mathrm{B}_{\sqrt{2}}(0)) \end{array}$

- If *x* is not the optimal solution, then $f(x)$ should decrease among the direction *d*:

$$f(x) < f(x+d) \Rightarrow \nabla f(x)^T d + f(x) < f(x) \tag{8}$$

$$\nabla f(x)^T d < 0 \tag{9}$$

- In order to have the inequality condition satisfied at *(x+d)*:

$$0 \ge g(x+d) \approx g(x) + \nabla g(x)^T d \tag{10}$$

$$0 \ge g(x) + \nabla g(x)^T d \tag{11}$$

For this example we have 2 cases: $x \in B_{\sqrt{2}}(0)$ and $x \in \partial B_{\sqrt{2}}(0)$

Case 1: $x \in B_{\sqrt{2}}(0)$

For this case it is clear that Eq. (11) is satisfied for sufficiently small *d* ($g(x) < 0$). For $\nabla f(x) \ne 0$, the direction $d \ne 0$ which satisfies both Eq. (9) and Eq. (11) is:

$$d = \alpha g(x) \frac{\nabla f(x)}{\|\nabla f(x)\|_2}, 0 < \alpha \begin{cases} +\infty & , \nabla g(x) = 0 \\ \dfrac{1}{\|\nabla g(x)\|_2} & , \nabla g(x) \ne 0 \end{cases} \tag{12}$$

In this case, the only way that makes the solution of *d* indefinite is the expression given in (13).

$$\nabla f(x) = 0 \tag{13}$$

Let's check if Eq.(12) is appropriate:

- *If* $\nabla g(x)=0 \Rightarrow$ <u>Eq. (11):</u> $g(x)+0\le 0$ *(according to ineq. cond.:* $g(x)<0$*)*

$$\text{\underline{Eq.(9):}}\ \nabla f(x)^T\left[\alpha . g(x).\frac{\nabla f(x)}{\|\nabla f(x)\|}\right]<0 \ \text{ because: } \begin{pmatrix} \nabla f(x)^T . \dfrac{\nabla f(x)}{\|\nabla f(x)\|_2}>0 \\ \alpha>0 \\ g(x)<0 \end{pmatrix}$$

If a vector is multiplied by itself, the answer is positive.
Here, any $\alpha>0$ *will do.*

- If $\nabla g(x)\neq 0 \Rightarrow$ Eq. (9) : *Again any* $\alpha>0$ *will be sufficient*

Eq. (11): if $\alpha=\dfrac{1}{\|\nabla g(x)\|_2}$ *then*

$$g(x)+g(x)\left[\frac{\nabla g(x)}{\|\nabla g(x)\|_2}.\frac{\nabla f(x)}{\|\nabla f(x)\|_2}\right]$$

Since the minimum value of the multiplication of two unit vectors is -1, we have

$$g(x)+g(x)\left[\frac{\nabla g(x)}{\|\nabla g(x)\|_2}.\frac{\nabla f(x)}{\|\nabla f(x)\|_2}\right]\ge g(x)-g(x)=0$$

Eq. (11) cannot be satisfied for any $\alpha>\dfrac{1}{\|\nabla g(x)\|_2}$ *.Oppositely* $\alpha\le\dfrac{1}{\|\nabla g(x)\|_2}$ *satisfies Eq. (11).*

<u>**Case 2:**</u> $x\in\partial B_{\sqrt{2}}(0)$

- If $g(x)=0$ then Eq. (9) and Eq. (11) will become $\nabla f(x)^T d<0$ and $\nabla g(x)^T d\le 0$ respectively. It is easy to see that these two conditions cannot be satisfied if $\nabla f(x)$ and $\nabla g(x)$ have opposite directions. So for a local minimum to exist at a certain point, we should have:

$$\nabla f(x)=-\mu\nabla g(x) \ \text{ for } \mu>0 \qquad (14)$$

For these two cases, the optimal condition can be written as Eq. (15).

$$L(x,\mu)=f(x)+\mu g(x) \qquad (15)$$

If $x^*\in F$ is a local minimum, then

$$\nabla_x L(x^*,\mu^*)=0 \qquad \textit{for } \mu^*>0 \qquad (15)$$

and

$$\mu^* g(x^*)=0 \qquad \text{(\underline{COMPLEMENTARY CONDITION})} \qquad (16)$$

Non-negative Matrix Factorization

For non-negative matrix factorization, the objective function is $f = tr\left[\mathbf{V}^T\mathbf{V} - 2\mathbf{V}^T\mathbf{WH} + \mathbf{H}^T\mathbf{W}^T\mathbf{WH}\right]$. The equality constraint is not valid while the inequality constraints are $g = -\mathbf{W} \le 0$ and $g = -\mathbf{H} \le 0$ (component-wise). This corresponds to "single inequality constraint" situation.

Since $\nabla_H f = \text{-}2\mathbf{W}^{\mathrm{T}}\mathbf{V}\text{+}2\mathbf{W}^{\mathrm{T}}\mathbf{WH}$ and $\frac{\partial g}{\partial \mathbf{H}} = \frac{\partial(-\mathbf{H})}{\partial \mathbf{H}} = \text{-}1$ we have the following expression for μ:

$$\nabla_H f = -\mu \, \nabla g \;\Rightarrow \text{-}2\mathbf{W}^{\mathrm{T}}\mathbf{V}\text{+}2\mathbf{W}^{\mathrm{T}}\mathbf{WH} = \text{-}\,\mu\,(\text{-}1)$$

When we use the expression of μ in the complementary KKT condition, we obtain:

$$\mu^* g\left(x^*\right) = 0 \Rightarrow \left(\text{-}2\mathbf{W}^{\mathrm{T}}\mathbf{V}\text{+}2\mathbf{W}^{\mathrm{T}}\mathbf{WH}\right)_{jk} G_{jk} = 0 \text{ (complementary KKT condition For } \mathbf{H}\text{)}$$

At convergence of the NMNF algorithm,

$$H^*_{kj} = H^*_{kj} \frac{\left(\mathbf{W}^{*^T}\mathbf{V}\right)_{\mathrm{kj}}}{\left(\mathbf{W}^{*^T}\mathbf{W}^*\mathbf{H}^*\right)_{\mathrm{kj}}}$$

$$H^*_{kj}\left(\mathbf{W}^{*^T}\mathbf{W}^*\mathbf{H}^*\right)_{\mathrm{kj}} = H^*_{kj}\left(\mathbf{W}^{*^T}\mathbf{V}\right)_{\mathrm{kj}}$$

$$H^*_{kj}\left(\mathbf{W}^{*^T}\mathbf{V} - \mathbf{W}^{*^T}\mathbf{W}^*\mathbf{H}^*\right)_{\mathrm{kj}} = 0$$

So, complementary KKT condition for **H** is satisfied. Similarly the proof can be done f**or W.**

References

[1] L.D. Berkowitz; Convexity and Optimization in Rn. Wiley-Interscience, New York, 2001.
[2] http://www.math.uh.edu/%7Erohop/fall_06/Chapter2.pdf