Steepest Descent Direction per Norm

Royi Avital

July 10, 2017

Contents

1 L_1 Norm **2**
1.1 Normalized Steepest Descent 2
1.2 Steepest Descent 3

2 L_2 Norm **3**
2.1 Normalized Steepest Descent 3
2.2 Steepest Descent 4

3 L_∞ Norm **4**
3.1 Normalized Steepest Descent 4
3.2 Steepest Descent 5

In the following derivations the vector f will represent the Gradient - $f = \nabla f(x)$. This is based on Computational Methods (046197) - Exercise 003.

Known related identity is given by:

$$d_{sd} = \|f\|_* d_{nsd}$$

Where d_{nds} is the Normalized Steepest Descent Direction with respect to the $\|\cdot\|$ norm.

1 L_1 Norm

1.1 Normalized Steepest Descent

$$\arg\min_{\|d\|_1 \leq 1} \left\{ f^T d \right\}$$

The Lagrangian of the problem can be written as:

$$L(d, \lambda) = f^T d + \lambda \left(\|d\|_1 - 1 \right)$$

The KKT Conditions are given by:

$$\begin{aligned}
\nabla_d L(d, \lambda) = f + \lambda \partial \|d\|_1 &= 0 && \text{(1) Stationary} \\
\lambda \left(\|d\|_1 - 1 \right) &= 0 && \text{(2) Slackness} \\
\|d\|_1 - 1 &\leq 0 && \text{(3) Primal Feasibility} \\
\lambda &\geq 0 && \text{(4) Dual Feasibility}
\end{aligned}$$

Looking at (1) yields:

$$-\frac{f}{\lambda} \in \partial \|d\|_1 = \begin{cases} 1 & \text{if } d_i > 0 \\ [-1, 1] & \text{if } d_i = 0 \\ -1 & \text{if } d_i < 0 \end{cases}$$

From above $\lambda > 0$ hence $\|d\|_1 = 1$. This is required as the problem is linear programming problem over a Polytope hence the solution must be on the boundaries of the set. Since $d \neq \mathbf{0}$ unless $f = \mathbf{0}$ then $\lambda = \max_i \{|f_i|\}$.This implies that:

$$d_i = \begin{cases} -\operatorname{sign}(f_i) & \text{if } |f_i| = \max_i \{|f_i|\} \\ 0 & \text{if } |f_i| \neq \max_i \{|f_i|\} \end{cases}$$

One could see the above minimizes one coordinate at a time which similar to Coordinate Descent.

1.2 Steepest Descent

$$\arg\min_{d}\left\{f^{T}d+\frac{1}{2}\|d\|_{1}^{2}\right\}$$

This is a constrains free convex problem hence the solution is given by a stationary point.

$$\begin{aligned}
0\in\partial\left(f^{T}d+\frac{1}{2}\|d\|_{1}^{2}\right)&=f+\|d\|_{1}\partial\|d\|_{1}\\
\Rightarrow-\frac{f}{\|d\|_{1}}&\in\begin{cases}1 & \text{if } d_i>0\\ [-1,1] & \text{if } d_i=0\\ -1 & \text{if } d_i<0\end{cases}
\end{aligned}$$

As above it means $\|d\|_{1}=\max_{i}\{|f_{i}|\}$ and:

$$d_{i}=\begin{cases}-f_{i} & \text{if } |f_{i}|=\max_{i}\{|f_{i}|\}\\ 0 & \text{if } |f_{i}|\neq\max_{i}\{|f_{i}|\}\end{cases}$$

Indeed $d_{sd}=\|f\|_{*}d_{sd}=\|f\|_{\infty}d_{sd}=\max_{i}\{|f_{i}|\}\,d_{sd}$ as required.

2 L_2 Norm

2.1 Normalized Steepest Descent

$$\arg\min_{\|d\|_{2}\leq 1}\left\{f^{T}d\right\}$$

The Lagrangian of the problem can be written as:

$$L(d,\lambda)=f^{T}d+\lambda\left(\|d\|_{2}-1\right)$$

The KKT Conditions are given by:

$$\begin{aligned}
\partial_{d}L(d,\lambda)=f+\lambda\partial\|d\|_{2}\ni 0 &\qquad \text{(1) Stationary}\\
\lambda\left(\|d\|_{2}-1\right)=0 &\qquad \text{(2) Slackness}\\
\|d\|_{2}-1\leq 0 &\qquad \text{(3) Primal Feasibility}\\
\lambda\geq 0 &\qquad \text{(4) Dual Feasibility}
\end{aligned}$$

Looking at (1) yields:

$$-\frac{f}{\lambda} \in \partial\|d\|_2 = \begin{cases} \frac{d}{\|d\|_2} & \text{if } d \neq 0 \\ \{g \mid \|g\|_2 \leq 1\} & \text{if } d = 0 \end{cases}$$

From above $\lambda > 0$ hence $\|d\|_2 = 1$ (Again, as expected by minimization of a linear function over a convex set). By choosing $\lambda = \|f\|_2$ and $d = -\frac{f}{\lambda} = -\frac{f}{\|f\|_2}$ all KKT conditions hold hence it is a feasible solution.

2.2 Steepest Descent

$$\arg\min_d \left\{ f^T d + \frac{1}{2}\|d\|_2^2 \right\}$$

This is a constrains free convex problem hence the solution is given by a stationary point.

$$\begin{aligned} 0 &= \nabla_d \left(f^T d + \frac{1}{2}\|d\|_2^2 \right) = f + d \\ &\Rightarrow d = -f \end{aligned}$$

Namely the solution, as expected, is the reflection (Negation) of the gradient $d = -f$ and $d_{sd} = \|f\|_* d_{sd} = \|f\|_2 d_{sd} = -\|f\|_2 \frac{f}{\|f\|_2} = -f$ as required.

3 L_∞ Norm

3.1 Normalized Steepest Descent

$$\arg\min_{\|d\|_\infty \leq 1} \left\{ f^T d \right\} = \arg\min_d \left\{ f^T d \right\} \text{ subject to } d \preceq 1, \, -d \preceq 1$$

The Lagrangian of the problem can be written as:

$$L(d, \lambda) = f^T d + \lambda_1^T (d - \mathbf{1}) + \lambda_2^T (-d - \mathbf{1})$$

The KKT Conditions are given by:

$$\begin{aligned} \nabla_d L(d, \lambda) = f + \lambda_1 - \lambda_2 &= 0 && \text{(1) Stationary} \\ \lambda_1^T (d - \mathbf{1}) &= 0 && \text{(2) Slackness} \\ \lambda_2^T (-d - \mathbf{1}) &= 0 && \text{(3) Slackness} \\ d - \mathbf{1} &\preceq 0 && \text{(4) Primal Feasibility} \\ -d - \mathbf{1} &\preceq 0 && \text{(5) Primal Feasibility} \\ \lambda_1, \lambda_2 &\succeq 0 && \text{(6) Dual Feasibility} \end{aligned}$$

The above can be solved component wise hence from now on the subscript for the component won't be written. Clearly λ_1 and λ_2 can not be both zero. Hence from (2) and (3) the components of d are either 1 or -1, namely $|d_i| = 1$. From (1) and (6) it must be $d_i = -\operatorname{sign}(f_i)$ hence $d = -\operatorname{sign}(f)$.

3.2 Steepest Descent

$$\arg\min_{d} \left\{ f^T d + \frac{1}{2}\|d\|_\infty^2 \right\}$$

This is a constrains free convex problem hence the solution is given by a stationary point.

$$\begin{aligned} 0 &\in \partial\left(f^T d + \frac{1}{2}\|d\|_\infty^2 \right) = f + \|d\|_\infty \partial\|d\|_\infty \\ \Rightarrow -\frac{f}{\|d\|_\infty} &\in \partial\|d\|_\infty = \left\{ w \mid \|w\|_1 \leq 1, w^T d = \|d\|_\infty \right\} \end{aligned}$$

In order to ensure $\left\| -\frac{f}{\|d\|_\infty} \right\|_1 \leq 1 \Rightarrow \|d\|_\infty \geq \|f\|_1$. From $-\frac{f^T}{\|d\|_\infty} d = \|d\|_\infty$ one could set $d = -\alpha \operatorname{sign}(f)$ which results in $\alpha f^T \operatorname{sign}(f) = \alpha \|f\|_1 = \|d\|_\infty^2$. Hence by setting $\alpha = \|f\|_1$ one would get $\|f\|_1 f^T \operatorname{sign}(f) = \|f\|_1^2 = \|d\|_\infty^2$ hence $d = -\|f\|_1 \operatorname{sign}(f)$.