Quadratic Programming

April 2, 2013

Introduction

Quadratic programming is the study on how to optimize a *multivariate* quadratic function.

It is very simple to find the optima of $f(x) = 2x^2 + 4x - 3$, but what about $f(x, y, z) = 2x^2 + 4y - 3z$?

1 Positive Definite matrix P and Convex Quadratic Function

Consider a quadratic function

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \mathbf{x} + 2\mathbf{q}^T \mathbf{x} + r$$

where $r \in \mathbb{R}^n$, $\mathbf{P} \in \mathbb{R}^n \times \mathbb{R}^n$

How to show $f(\mathbf{x})$ is convex iff $\mathbf{P}$ is positive semidefinite ?

Recall ,

Function is convex iff

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \qquad \forall \alpha \in [0,1]$$

$\mathbf{P}$ is positive semidefinite iff

$$\mathbf{x}^T \mathbf{P} \mathbf{x} \geq 0$$

Let do the proof by

I. Assuming $\mathbf{P}$ is positive semidefinite, to show $f(\mathbf{x})$ is convex by
1. Expand $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})$
2. Apply the assumption $\mathbf{P}$ is positive semidefinite
3. Show $f(\mathbf{x})$ is convex by $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$

II. Assume $f(\mathbf{x})$ is convex, show $\mathbf{P}$ is positive semidefinite (skiped here, the method is the same)

1. Since $f(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r$, thus replace $\mathbf{x}$ to $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) = [\alpha\mathbf{x} + (1-\alpha)\mathbf{y}]^T \mathbf{P} [\alpha\mathbf{x} + (1-\alpha)\mathbf{y}] + 2\mathbf{q}^T [\alpha\mathbf{x} + (1-\alpha)\mathbf{y}] + r$$

Expand it step-by-step

$$= \left[\alpha\mathbf{x}^T + (1-\alpha)\mathbf{y}^T\right] [\alpha\mathbf{P}\mathbf{x} + (1-\alpha)\mathbf{P}\mathbf{y}] + 2\alpha\mathbf{q}^T\mathbf{x} + 2(1-\alpha)\mathbf{q}^T\mathbf{y} + r$$

$$= \left[\alpha^2\mathbf{x}^T\mathbf{P}\mathbf{x} + \alpha(1-\alpha)\mathbf{y}^T\mathbf{P}\mathbf{x} + \alpha(1-\alpha)\mathbf{x}^T\mathbf{P}\mathbf{y} + (1-\alpha)^2\mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y}\right] + 2\alpha\mathbf{q}^T\mathbf{x} + 2(1-\alpha)\mathbf{q}^T\mathbf{y} + r$$

Group those terms with same α terms

$$= \alpha^2\mathbf{x}^T\mathbf{P}\mathbf{x} + \alpha(1-\alpha)\left[\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y}\right] + (1-\alpha)^2\mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y} + 2\alpha\mathbf{q}^T\mathbf{x} + 2(1-\alpha)\mathbf{q}^T\mathbf{y} + r$$

The following step is a little bit mathematics,

Consider

$$1 = (1-\alpha) + \alpha \qquad \alpha^2 = \alpha - \alpha(1-\alpha) \qquad (1-\alpha)^2 = ((1-\alpha) - \alpha(1-\alpha))$$

Thus

$$\begin{aligned}&= \{\alpha - \alpha(1-\alpha)\}\, \mathbf{x}^T\mathbf{P}\mathbf{x} \\ &\quad + \alpha(1-\alpha)\left[\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y}\right] + \{(1-\alpha) - \alpha(1-\alpha)\}\, \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y} + 2\alpha\mathbf{q}^T\mathbf{x} + 2(1-\alpha)\mathbf{q}^T\mathbf{y} + \{(1-\alpha) + \alpha\}\, r\end{aligned}$$

Group the terms with same α terms, thus

$$= \alpha\left[\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r\right] + (1-\alpha)\left[\mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y} + 2\mathbf{q}^T\mathbf{y} + r\right] + \alpha(1-\alpha)\left[\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y} - \mathbf{x}^T\mathbf{P}\mathbf{x} - \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y}\right]$$

2. Next, consider $\left[\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y} - \mathbf{x}^T\mathbf{P}\mathbf{x} - \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y}\right]$, and apply the assumption of $\mathbf{P}$ is positive semidefinite

$$\begin{aligned}\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y} - \mathbf{x}^T\mathbf{P}\mathbf{x} - \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y} &= \mathbf{y}^T\mathbf{P}\mathbf{x} - \mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y} - \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y} \\ &= \left(\mathbf{y}^T - \mathbf{x}^T\right)\mathbf{P}\mathbf{x} + \left(\mathbf{x}^T - \mathbf{y}^\mathbf{T}\right)\mathbf{P}\mathbf{y} \\ &= \left(\mathbf{y}^T - \mathbf{x}^T\right)\mathbf{P}\mathbf{x} - \left(\mathbf{y}^\mathbf{T} - \mathbf{x}^T\right)\mathbf{P}\mathbf{y} \\ &= \left(\mathbf{y}^T - \mathbf{x}^T\right)\mathbf{P}\left(\mathbf{x} - \mathbf{y}\right) \\ &= \left(\mathbf{y} - \mathbf{x}\right)^T\mathbf{P}\left(\mathbf{x} - \mathbf{y}\right) \\ &= -\left(\mathbf{y} - \mathbf{x}\right)^T\mathbf{P}\left(\mathbf{y} - \mathbf{x}\right) \\ &\le 0\end{aligned}$$

3. Consider $\alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$

$$= \alpha\left(\mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r\right) + (1-\alpha)\left(\mathbf{y}^T\mathbf{P}\mathbf{y} + 2\mathbf{q}^T\mathbf{y} + r\right)$$

which is the first 2 term in $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})$!

Thus

$$\begin{aligned}
f(\mathbf{x}) &= \mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r \\
&= \alpha\left[\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r\right] + (1-\alpha)\left[\mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y} + 2\mathbf{q}^T\mathbf{y} + r\right] + \alpha(1-\alpha)\left[\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y} - \mathbf{x}^T\mathbf{P}\mathbf{x} - \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y}\right] \\
&= \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) + \alpha(1-\alpha)\left[\mathbf{y}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\mathbf{y} - \mathbf{x}^T\mathbf{P}\mathbf{x} - \mathbf{y}^\mathbf{T}\mathbf{P}\mathbf{y}\right] \\
&\leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})
\end{aligned}$$

The above inequality holds as $\alpha(1-\alpha) \geq 0$, since $\alpha \in [0, 1]$

Therefore, $f(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r$ is convex if $\mathbf{P}$ is semidefinite

2 Minimizing a quadratic function

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r$$

If $\mathbf{P}$ is positive semidefinite, then f is convex , and thus the $\mathbf{x}*$ is the optimal solution.

How to find $\mathbf{x}*$?

Using the first-order condition,

$$\nabla f(\mathbf{x}*) = \nabla(\mathbf{x}*^T\mathbf{P}\mathbf{x}* + 2\mathbf{q}^T\mathbf{x}* + r) = 0$$

Recall *vector and matrix differentiation*

$$\frac{d}{d\mathbf{x}} \hat{=} \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{bmatrix} \qquad \frac{df(\mathbf{x})}{d\mathbf{x}} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

$$\frac{d}{d\mathbf{x}}\left(\mathbf{x}^\mathbf{T}\mathbf{x}\right) = 2\mathbf{x} \qquad \frac{d}{d\mathbf{x}}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = 2\mathbf{A}\mathbf{x} \quad \text{if A is symmetric}$$

Simple proof

$$\frac{d}{d\mathbf{x}}\left(\mathbf{x}^\mathbf{T}\mathbf{x}\right) = \frac{d}{d\mathbf{x}}\left(x_1^2 + x_2^2 + ... + x_n^2\right) = \begin{bmatrix} 2x_1 \\ \vdots \\ 2x_n \end{bmatrix} = 2\mathbf{x}$$

Thus

$$\nabla f(\mathbf{x}*) = 2\mathbf{P}\mathbf{x}* + 2\mathbf{q}^T = 0$$

i.e.

$$\mathbf{x}* = -\mathbf{P}^{-1}\mathbf{q}^\mathbf{T}$$

is a global minizier of $f(\mathbf{x})$

3 Linear Least-Square Estimation : The Wiener Filter

Let $\mathbf{a} \in \mathbb{R}^{\mathbf{n}}$ be input to a linear system that output $y \in \mathbb{R}$ is given by $y = \mathbf{a}^T\mathbf{x}$

To estimate $\mathbf{x}$, which is a constant vector, just apply different $\mathbf{a}_i$ to the system and measure the corresponding output y_i

But real world is full of noise n_i , thus

$$y_i = \mathbf{a}_i^T\mathbf{x} + n_i$$

Thus the best way to estimate $\mathbf{x}$ is to minimizing energy of n_i

In matrix form

$$\begin{matrix} y_1 = \mathbf{a_1^T}\mathbf{x} + n_1 \\ \vdots \\ y_n = \mathbf{a_n^T}\mathbf{x} + n_n \end{matrix} \quad \Longleftrightarrow \quad \mathbf{y} = \mathbf{Ax} + \mathbf{n} \quad \begin{cases} \mathbf{y} = [y_1 \,, \, ... \,, \, y_n]^T \\ \mathbf{A} = [\mathbf{a_1} \,, \, ... \,, \, \mathbf{a_n}]^T \\ \mathbf{n} = [n_1 \,, \, ... \,, \, n_n]^T \end{cases}$$

Thus the problem can be wirtten as

$$\min_{x \in \mathbb{R}^n} ||\mathbf{y} - \mathbf{Ax}||_2^2 = \min_{x \in \mathbb{R}^n} ||\mathbf{n}||_2^2$$

Thus

$$\begin{aligned} \min_{x \in \mathbb{R}^n} ||\mathbf{y} - \mathbf{Ax}||_2^2 &= \min_{x \in \mathbb{R}^n} (\mathbf{y} - \mathbf{Ax})^T (\mathbf{y} - \mathbf{Ax}) \\ &= \min_{x \in \mathbb{R}^n} \left(\mathbf{y^T} - \mathbf{x^T A^T}\right)(\mathbf{y} - \mathbf{Ax}) \\ &= \min_{x \in \mathbb{R}^n} \mathbf{y^T y} - \mathbf{y^T A x} - \mathbf{x^T A^T y} + \mathbf{x^T A^T A x} \\ &= \min_{x \in \mathbb{R}^n} \mathbf{y^T y} - \mathbf{y^T A x} - \mathbf{x^T A^T y} + \mathbf{x^T A^T A x} \\ &= \min_{x \in \mathbb{R}^n} \mathbf{y^T y} - \mathbf{y^T A x} - (\mathbf{Ax})^{\mathbf{T}} \mathbf{y} + \mathbf{x^T A^T A x} \\ &= \min_{x \in \mathbb{R}^n} \mathbf{y^T y} - \mathbf{2 y^T A x} + \mathbf{x^T A^T A x} \end{aligned}$$

Note 1 . Recall $(AB)^T = B^T A^T$ and $(A+B)^T = A^T + B^T$
Note 2.$(\mathbf{Ax})^{\mathbf{T}}\mathbf{y} = \mathbf{y^T A x}$ because the output is a scalar

Thus

$$\min_{x \in \mathbb{R}^n} ||\mathbf{y} - \mathbf{Ax}||_2^2 = \min_{x \in \mathbb{R}^n} \mathbf{y^T y} - \mathbf{2 y^T A x} + \mathbf{x^T A^T A x}$$

Compare to quadratic function

$$\mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r$$

$$||\mathbf{y} - \mathbf{Ax}||_{\mathbf{2}}^{\mathbf{2}} = \mathbf{x^T} \underbrace{(A^T A)}_{\mathbf{P}} \mathbf{x} - \mathbf{2} \underbrace{y^T A}_{-\mathbf{q}} \mathbf{x} + \underbrace{\mathbf{y^T y}}_{r}$$

Assume $\mathbf{P} = \mathbf{A^T A}$, called *normal matrix* , is invertible

And $\mathbf{y^T A} = \left(\mathbf{A^T y}\right)^{\mathbf{T}} = -\mathbf{q}$, that is $\mathbf{q} = -\mathbf{A^T y}$

Then when $\mathbf{x}* = -\mathbf{P^{-1} q^T} = \mathbf{P^{-1} A^T y}$, the solution will be optimal, and this $\mathbf{x}*$is called *Wiener Filter*

$-E_{N_D}-$