Dati N punti $\{(x_i, y_i)\}$, $y_i = f(x_i)$, $1 \le i \le N$ e $m < N$, il vettore $a \in \mathbb{R}^{m+1}$ minimizza $\phi(a) = \sum_{i=1}^{N} (y_i - \sum_{j=0}^{m} a_j \cdot x_i^j)^2 \Longleftrightarrow$ risolve il sistema $V^t V a = V^t y$

Dire che $a \in \mathbb{R}^{m+1}$ e' di minimo assoluto per $\phi(a)$ equivale a dire:

$\phi(a+b) \ge \phi(a) \;\forall\, b \in \mathbb{R}^{m+1}$ ma $\phi(a+b) = \phi(a) + 2(V^t(Va - y), b) + (Vb, Vb)$

Dim "$\Leftarrow$": assumendo che $V^t V a = V^t y$ abbiamo:

$V^t V a - V^t y = V^t (Va - y) = 0$ e $(V^t(Vb - y), b) = (0, b)$ vettore nullo in $\mathbb{R}^{m+1} = 0$

Dim "$\Rightarrow$": assumendo che $\phi(a+b) \ge \phi(a) \;\forall\, b \in \mathbb{R}^{m+1}$ allora:

$\phi(a+b) = \phi(a) + 2(V^t(Vb - y), b) + (Vb, Vb) \ge \phi(a) \;\forall\, b$ cioè:

$2(V^t(Vb - y), b) + (Vb, Vb) \ge 0 \;\forall\, b.$

Ponendo $b = \varepsilon v$, con v vettore versore $((v,v) = 1)$. Dividendo per $\varepsilon > 0$ e considerando $\varepsilon \to$ otteniamo:

$(V^t(Va - y), v) \ge 0 \;\forall v$

Ma essendo la disuguaglianza valida per ogni v, allora possiamo sostituire $-v$ a v e otteniamo:

$(V^t(Va - y), -v) = -(V^t(Va - y), v) \ge 0 \;\forall v$ cioè

$0 \le (V^t(Va - y), v) \le 0 \Rightarrow (V^t(Va - y), v) = 0 \;\forall v$

Ma essendo il vettore nullo l'unico vettore ortogonale a tutti i vettori:

$V^t(Va - y) = 0 \Longleftrightarrow V^t V a = V^t y$ ovvero "a" soluzione del sistema