Métodos de otimização mais eficientes

Otimização via Métodos Numéricos Recursivos

$$\underline{w} \rightarrow \underline{w} + \underline{\Delta w} \quad | \quad F(\underline{w} + \underline{\Delta w}) < F(\underline{w})$$

$\underline{\Delta w}$??

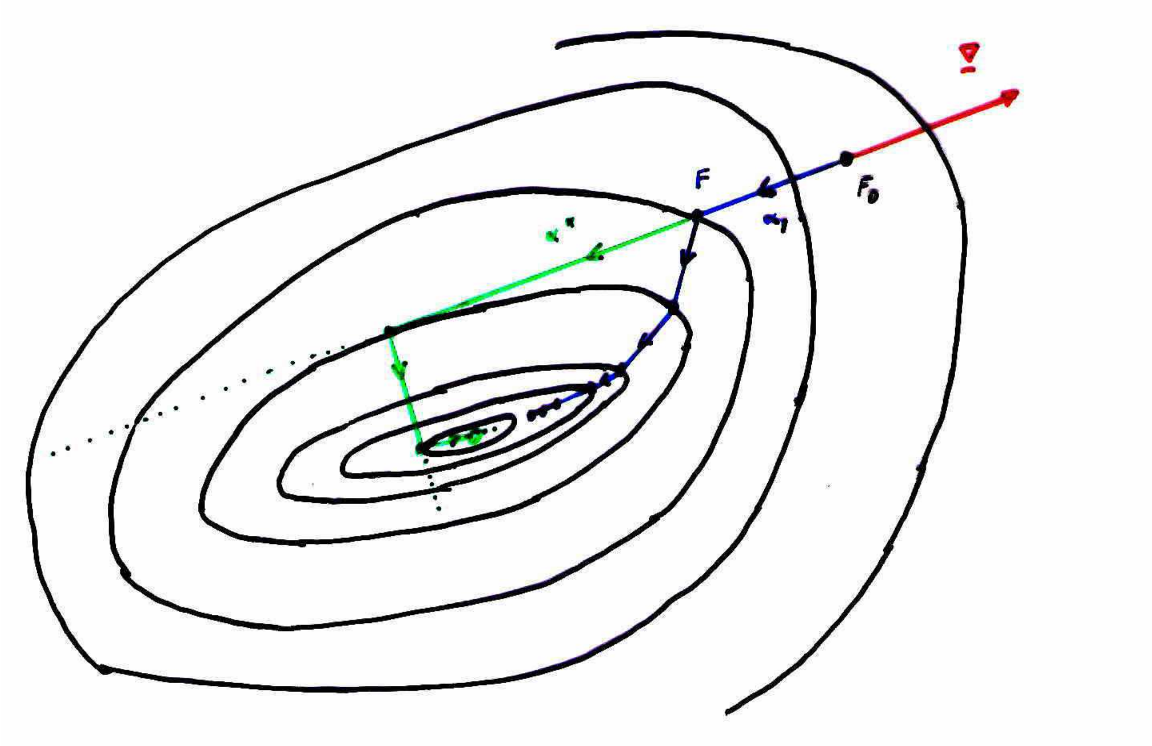
$$\underline{\Delta w} = \alpha \, \underline{d}$$

Passo
controla | $\underline{\Delta w}$ |
$\alpha > 0$, pequeno

controla
direção $\underline{\Delta w}$

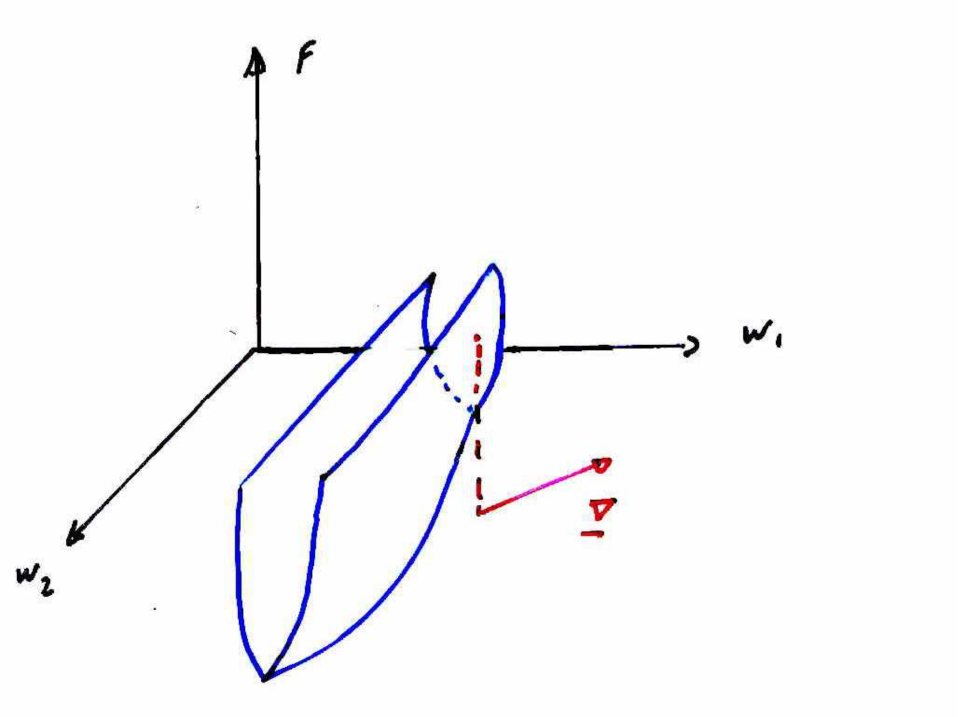
1 - Passo variável, otimizado

Passo ótimo ? otimização em linha



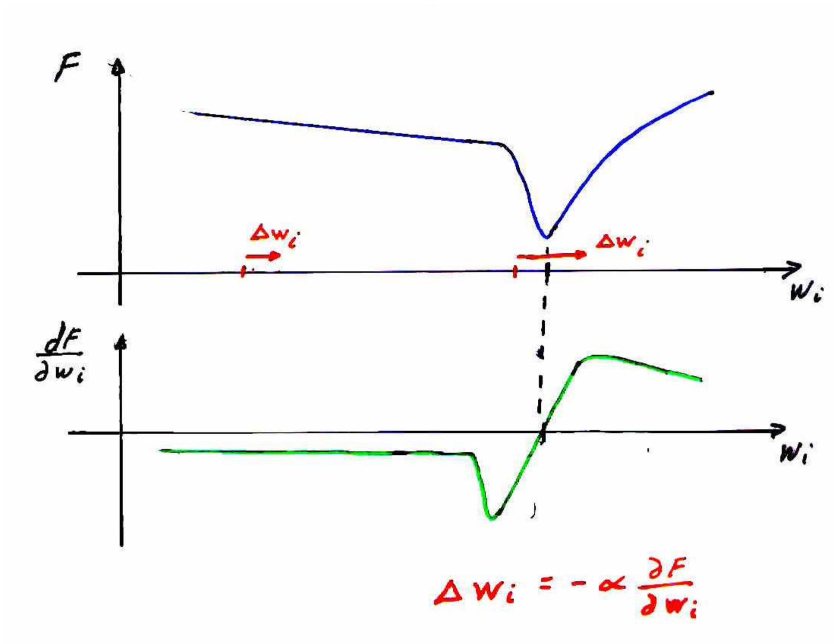
Passo variável por sinapse

α diferenciado por sinapse



Passo variável por ponto de operação

α diferenciado por passo de treinamento



1.1 - Resilient Backpropagation (modificado)

(Silva e Almeida, Super SAB, etc.)

$$\alpha \text{ variável } \begin{cases} \text{por variável } w_i \\ \text{por passo de treinamento } n \end{cases}$$

$$\Delta w_i(n) = -\,\alpha_i(n)\,\frac{\partial F}{\partial w_i}(n)$$

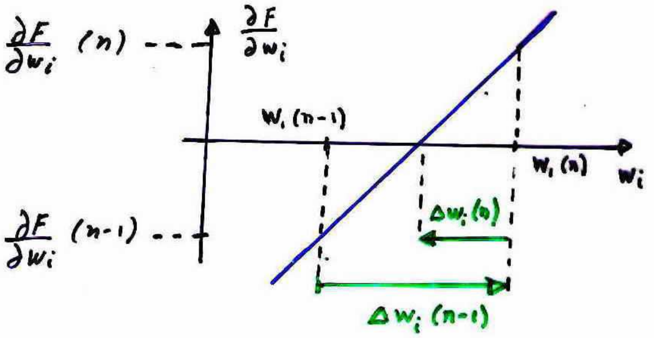
$$\alpha_i(n) = \begin{cases} a\,\alpha_i(n-1) & \text{se } \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n) = \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n-1) \\ b\,\alpha_i(n-1) & \text{se } \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n) \neq \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n-1) \end{cases}$$

se $\alpha_i(n) > \alpha^*$ faça $\alpha = \alpha^*$

$\alpha_i(0) \approx .05$ $\alpha^* \approx .2$ $a \approx 1.05$ $b \approx .9$

Outra alternativa

$$b = \frac{\dfrac{\partial F}{\partial w_i}(n-1)}{\dfrac{\partial F}{\partial w_i}(n-1) - \dfrac{\partial F}{\partial w_i}(n)}$$



BP Resiliente - Algorítmo

Inicialização

$n = 1 \quad \alpha_{\max} \approx .5 \quad a \approx 1.05$

para as i = 1, 2, …, V variáveis

$w_i(1) \in [+.2, -.2]$ randômicos; $\alpha_i(1) = .1 \quad \frac{\partial F}{\partial w_i}(0) = 0$

Até que o critério de parada esteja satisfeito

Passo de treinamento n

BP Resiliente - Algorítmo (continuação)

Cálculo do gradiente

para cada par ($\underline{x}^p$, $\underline{y}^p$) p = 1, …, P

calcular $\tilde{\underline{y}}^p = \varphi\left(\underline{x}^p\right)$

$\varepsilon_m^p = y_m^p - \tilde{y}_m^p \quad$ e $\quad \frac{\partial \tilde{y}_m^p}{\partial w_i} \quad \forall\, m, i$

$$\frac{\partial \varepsilon^{p2}}{\partial w_i} = -2 \sum_{m=1}^{M} \varepsilon_m^p \frac{\partial y_m^p}{\partial w_i} \quad \forall\, i$$

outro par

calcular $\frac{\partial F}{\partial w_i}(n) = \underset{p}{E} \frac{\partial \varepsilon^{p2}}{\partial w_i}$

BP Resiliente - Algorítmo (continuação)

Cálculo do acréscimo

para i = 1, 2, …, V

$$\alpha_i(n) = \begin{cases} a\,\alpha_i(n-1) & \text{se } \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n) = \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n-1) \\ & \textit{mas}\text{ se } \alpha_i(n) > \alpha_{max}\ \textit{faça}\ \alpha_i(n) = \alpha_{max} \\ \dfrac{\dfrac{\partial F}{\partial w_i}(n-1)}{\dfrac{\partial F}{\partial w_i}(n-1) - \dfrac{\partial F}{\partial w_i}(n)}\,\alpha_i(n-1) & \text{se } \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n) \neq \operatorname{sign}\dfrac{\partial F}{\partial w_i}(n-1) \end{cases}$$

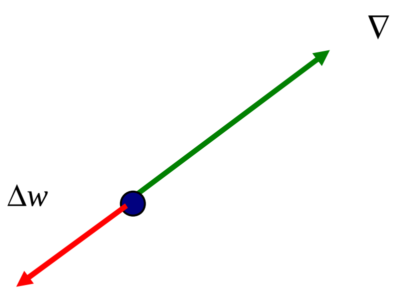
$$\Delta w_i(n) = -\,\alpha_i(n)\,\frac{\partial F}{\partial w_i}(n)$$

$$w_i(n+1) = w_i(n) + \Delta w_i(n)$$

n = n + 1

1.2 - Métodos de 1ª. Ordem (Resumo)

$$\underline{w}_{n+1} = \underline{w}_n + \alpha_n\,\underline{d}_n$$



$$\underline{d}_n^t = -\,\underline{\nabla}_n$$

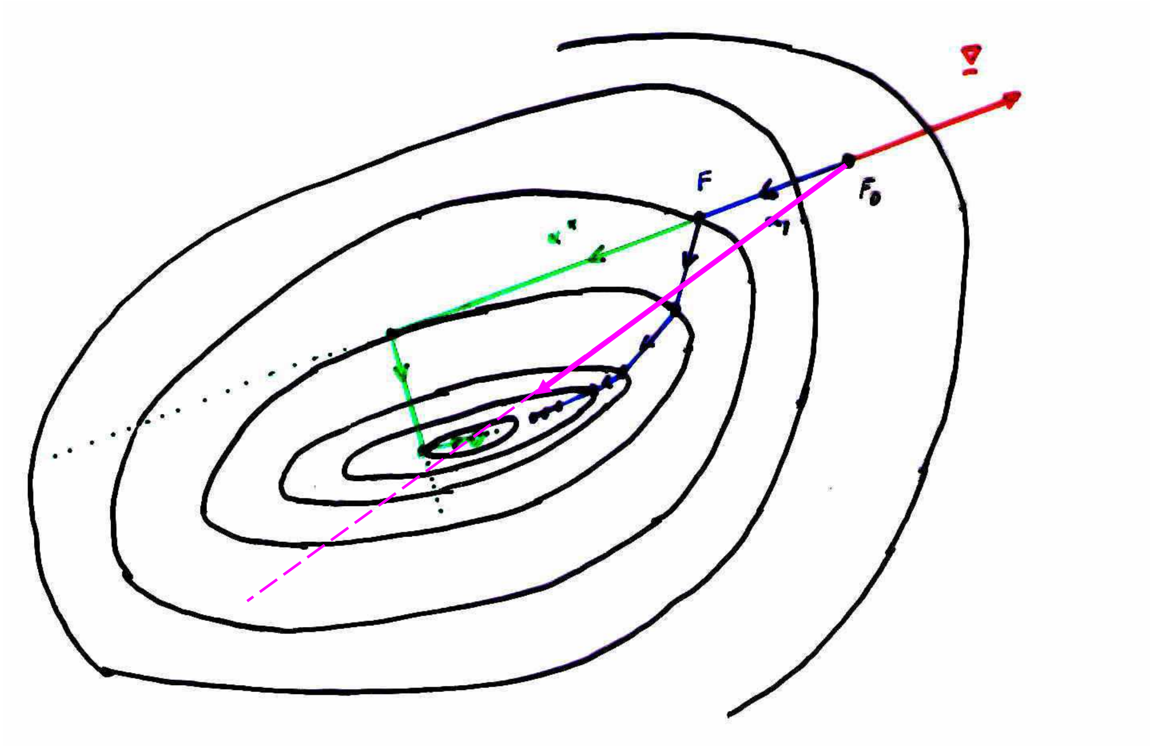
BP : α constante

BP Resiliente: α variável por sinapse e por passo

2 - Métodos de Segunda Ordem

Passo variável, otimizado <u>e</u>

Melhores direções de deslocamento



Métodos de primeira ordem dependem de

$$\underline{\Delta w} = \left.\Delta w_i\right] \qquad \nabla F(\underline{w}_0) = \left.\frac{\partial F}{\partial w_i}\right|_{\underline{w}_0} \Bigg]$$ **gradiente**

Métodos de segunda ordem dependem também de

$$\underline{H}(\underline{w}_0) = \left[\left.\frac{\partial^2 F}{\partial w_i \, \partial w_j}\right|_{\underline{w}_0} \right]$$ **Hessiana**

ou de suas aproximações

MÉTODOS DE SEGUNDA ORDEM:
MAIS SOFISTICADOS / MAIS EFICIENTES (?)

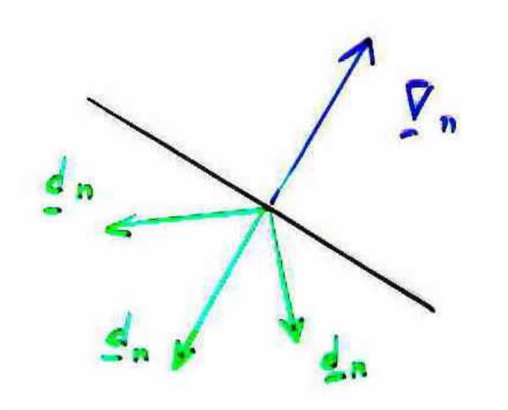
Newton	$\underline{\mathbf{H}}$	
Newton Amortecido	$\approx \underline{\mathbf{H}}$	Levenberg – Marquadt
Quasi Newton	$\approx \underline{H}^{-1}$	BFGS
Gradiente Conjugado		Fletcher – Reeves

Métodos de 2ª. Ordem (Resumo)

$$\underline{w}_{n+1} = \underline{w}_n + \alpha_n \underline{d}_n$$

$$\alpha_n = \underset{\alpha_n > 0}{\mathrm{Arg}} \left[\mathrm{Min}\ F(\underline{w}_n + \alpha_n \underline{d}) \right] \quad \text{otimização em linha}$$

$\underline{d}_n$ **???**



Referências bibliográficas principais:

Cichocki, Unbehauen – “Neural Networks for Optimization and Signal Processing”, Cap. 1 e 3, Wiley, 1993.

Chong, Zak – “An Introduction to Optimization”, Part II, Wiley, 2001.

Addy, Dempster – “Introduction to optimization methods”, Chapman and Hall, 1974.