Topics on nonlinear optimization

Contents

- Introduction

- Gradient based methods
  - Newton's method
  - Gauss-Newton's method
  - Steepest Decent
  - Levenberg-Marquardt's method

Introduction

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \quad \mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}$$

$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\nabla \Phi(\mathbf{p}^*) = \mathbf{0}_{M \times 1}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}$$

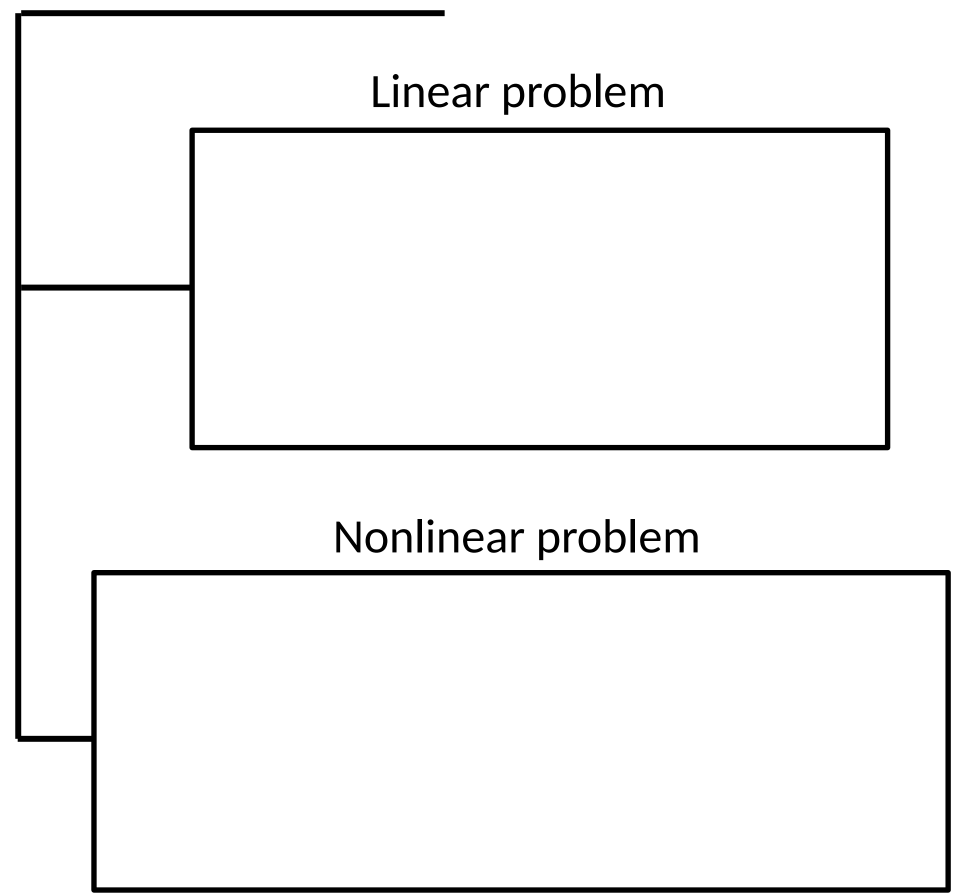
$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}$$

$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\nabla\Phi(\mathbf{p}^*) = \mathbf{0}_{M\times 1}$$



$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \boxed{\mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}}$$

$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\nabla \Phi(\mathbf{p}^*) = \mathbf{0}_{M \times 1}$$

Linear problem

Nonlinear problem

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \boxed{\mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}}$$

$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\nabla\Phi(\mathbf{p}^*) = \mathbf{0}_{M\times 1}$$

Linear problem

$$\mathbf{g}(\mathbf{p}) = \mathbf{B}\mathbf{p} + \mathbf{b}$$

$$\mathbf{p}^* = \left(\mathbf{B}^\top\mathbf{B}\right)^{-1}\mathbf{B}^\top(\mathbf{d} - \mathbf{b})$$

Nonlinear problem

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \boxed{\mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}}$$

$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\nabla\Phi(\mathbf{p}^*) = \mathbf{0}_{M\times 1}$$

Linear problem

$$\mathbf{g}(\mathbf{p}) = \mathbf{B}\mathbf{p} + \mathbf{b}$$

$$\mathbf{p}^* = \left(\mathbf{B}^\top\mathbf{B}\right)^{-1}\mathbf{B}^\top(\mathbf{d} - \mathbf{b})$$

Nonlinear problem

$$\mathbf{g}(\mathbf{p}) \neq \mathbf{B}\mathbf{p} + \mathbf{b}$$

$$\mathbf{p}_k = \mathbf{p}_{k-1} + \Delta\mathbf{p}_k$$

$$\mathbf{p}^* \approx \mathbf{p}_0 + \Delta\mathbf{p}_1 + \cdots + \Delta\mathbf{p}_L$$

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \qquad \boxed{\mathbf{g}(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ g_2(\mathbf{p}) \\ \vdots \\ g_N(\mathbf{p}) \end{bmatrix}}$$

$$\Phi(\mathbf{p}) = [\mathbf{d} - \mathbf{g}(\mathbf{p})]^\top [\mathbf{d} - \mathbf{g}(\mathbf{p})]$$

$$\nabla\Phi(\mathbf{p}^*) = \mathbf{0}_{M\times 1}$$

Linear problem

$$\mathbf{g}(\mathbf{p}) = \mathbf{B}\mathbf{p} + \mathbf{b}$$

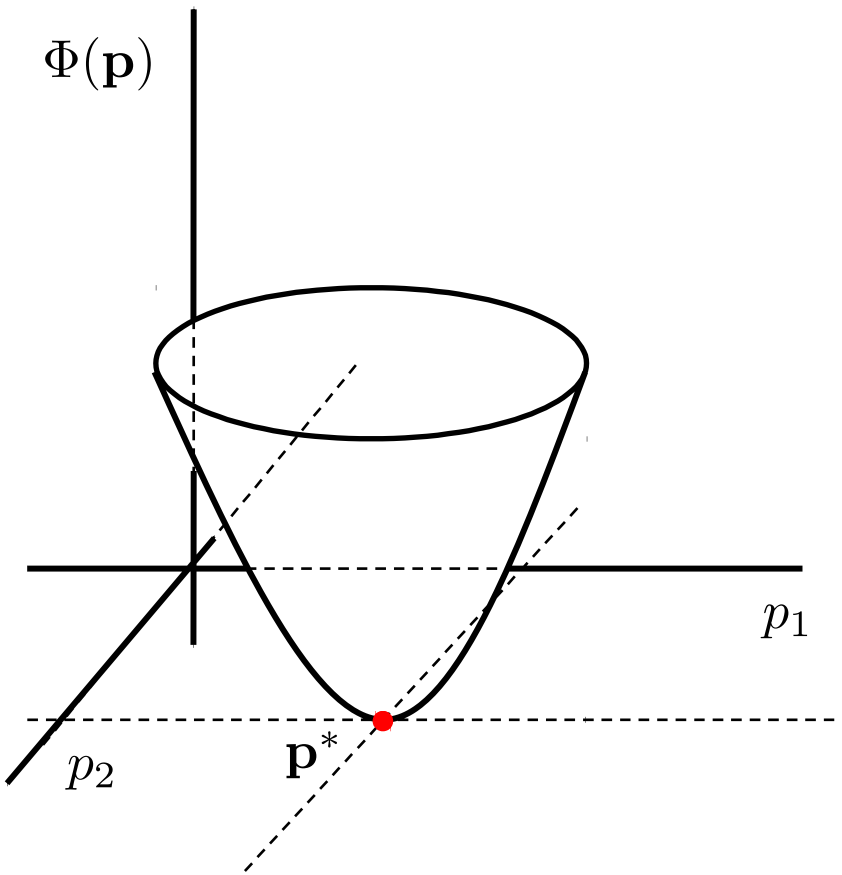
$$\mathbf{p}^* = \left(\mathbf{B}^\top\mathbf{B}\right)^{-1}\mathbf{B}^\top(\mathbf{d} - \mathbf{b})$$

Nonlinear problem

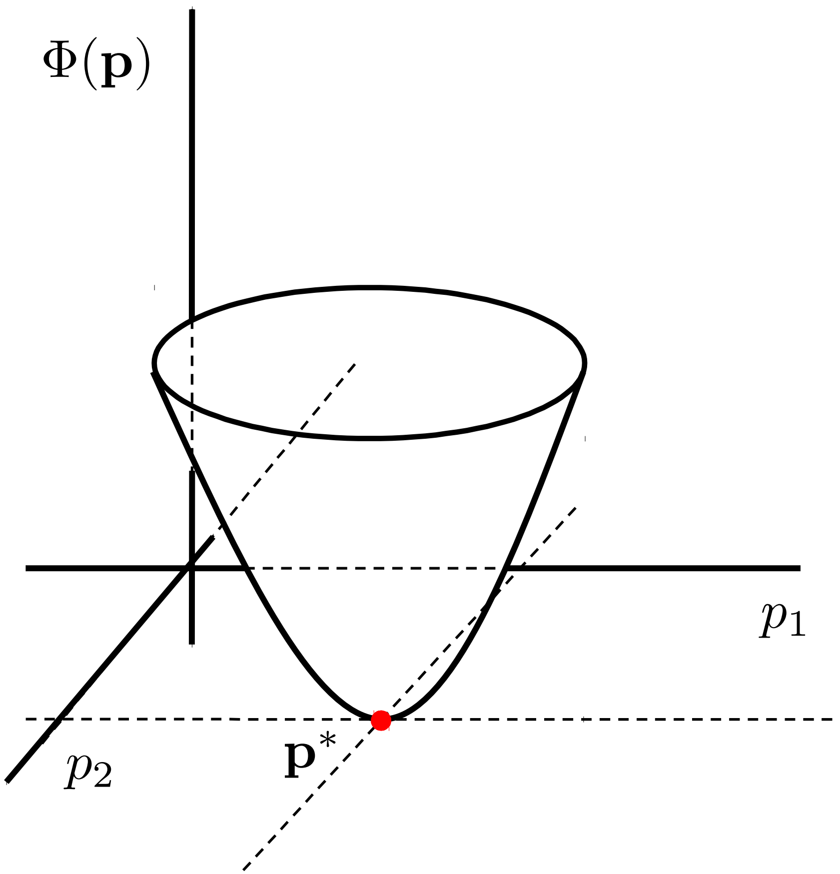
$$\mathbf{g}(\mathbf{p}) \neq \mathbf{B}\mathbf{p} + \mathbf{b}$$

$$\mathbf{p}_k = \mathbf{p}_{k-1} + \Delta\mathbf{p}_k$$

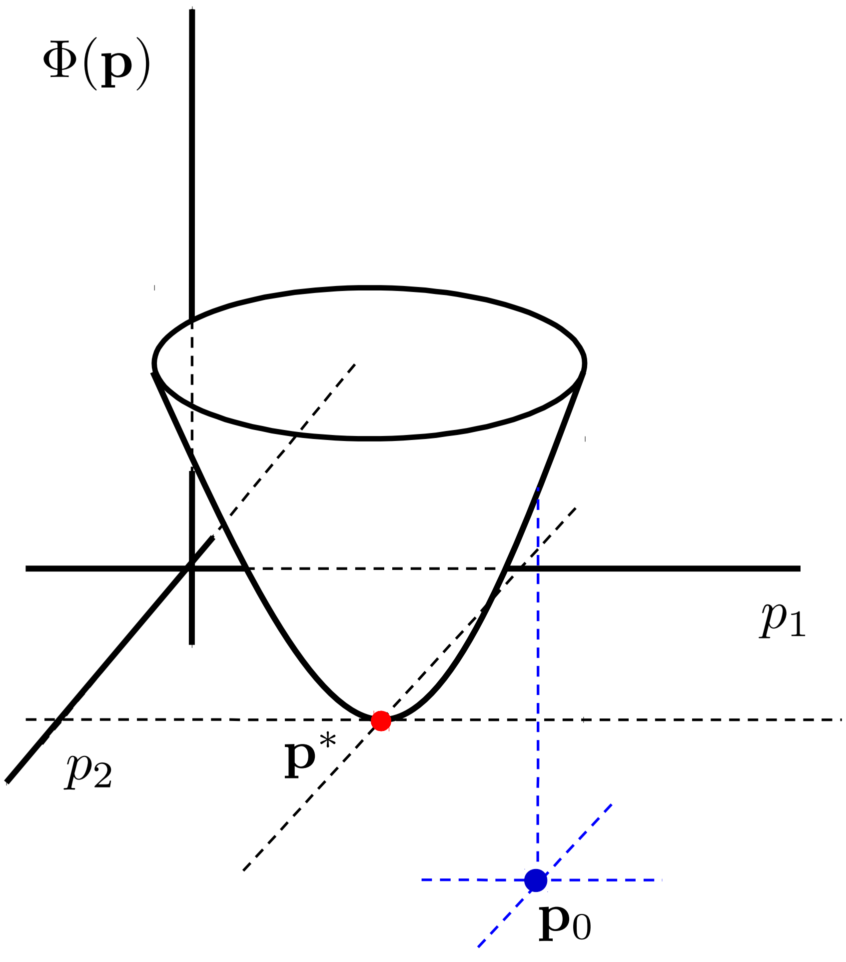
$$\mathbf{p}^* \approx \mathbf{p}_0 + \Delta\mathbf{p}_1 + \cdots + \Delta\mathbf{p}_L$$



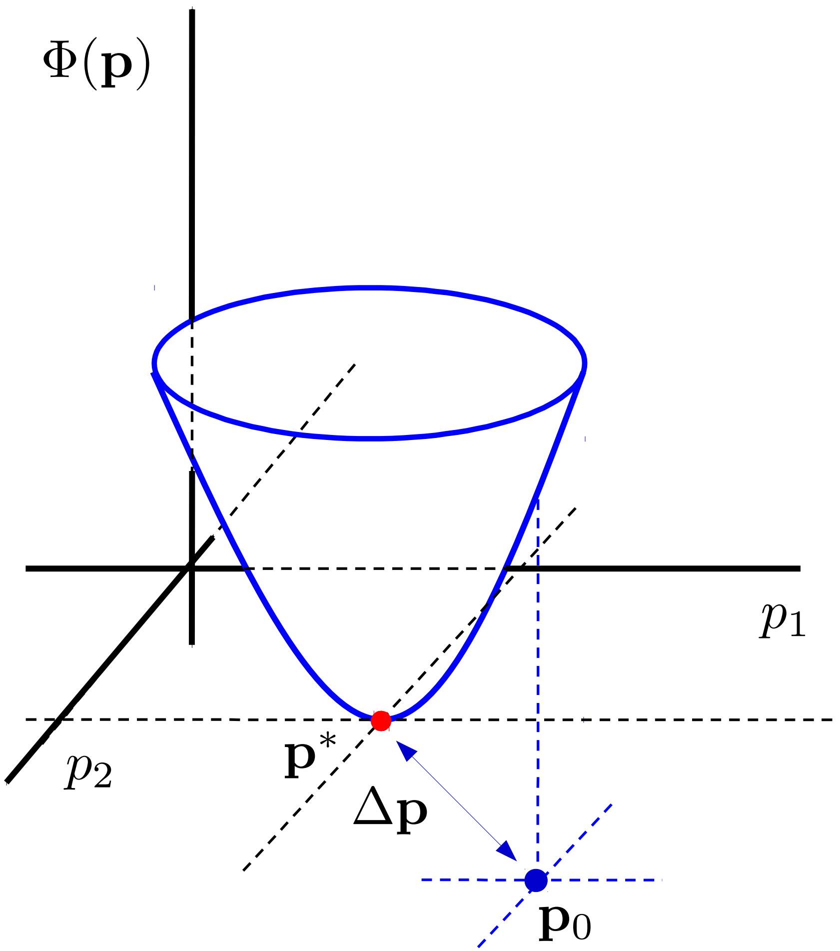
Linear problem



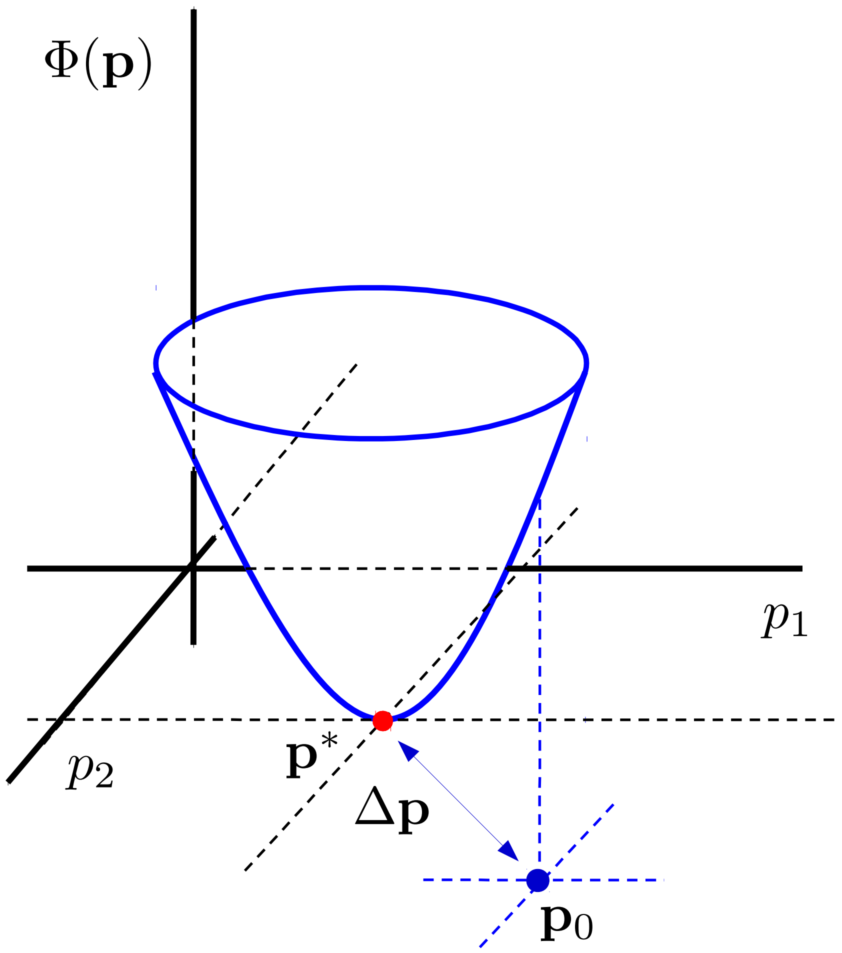
The minimum can be computed in a single step



Or iteratively, from a given initial approximation

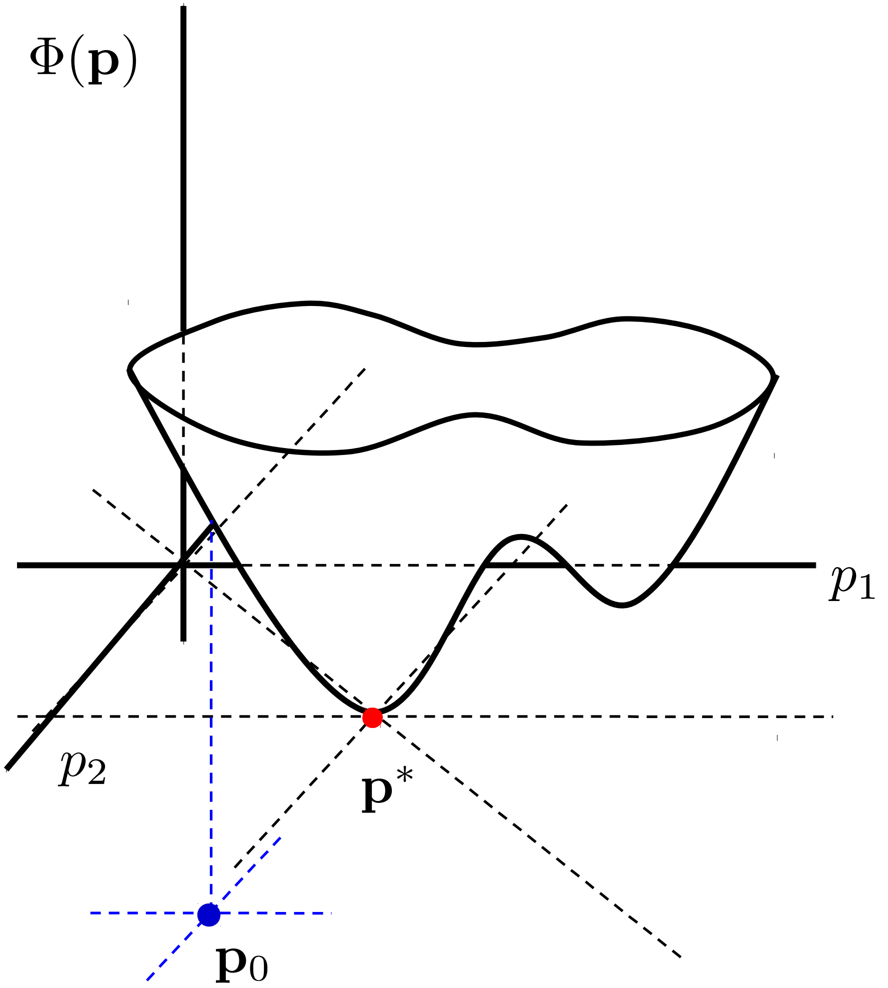


$\Phi(\mathbf{p})$
p_1
p_2
$\mathbf{p}^*$
$\Delta\mathbf{p}$
$\mathbf{p}_0$

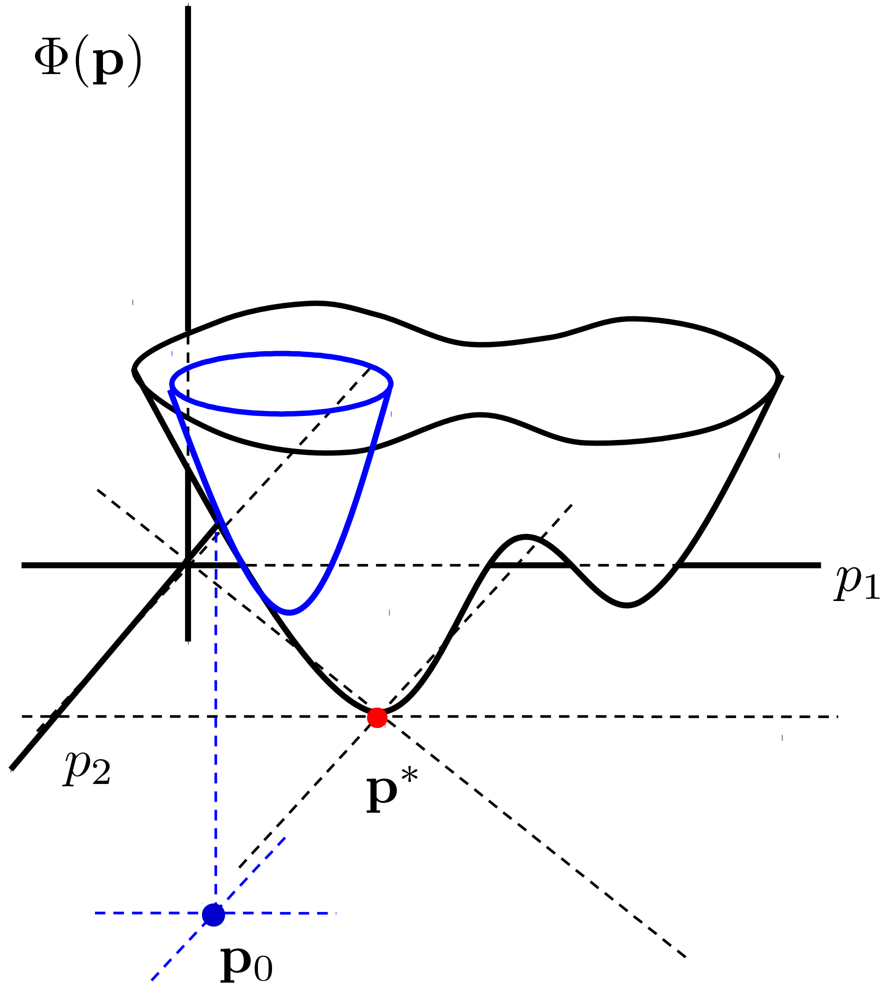


In this case, the minimum is estimated in a single step from the given initial approximation

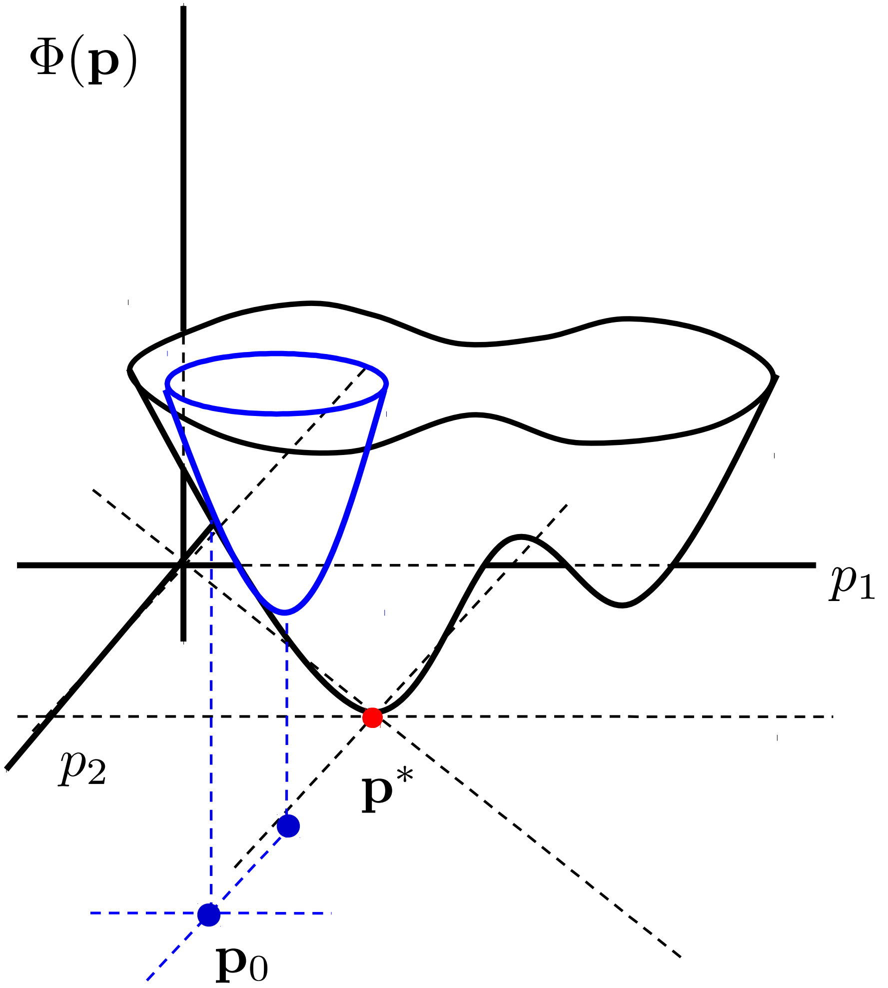
On the other hand, in a nonlinear problem, the minimum is estimated after several steps from the initial approximation



Approximate the nonlinear function around the initial approximation

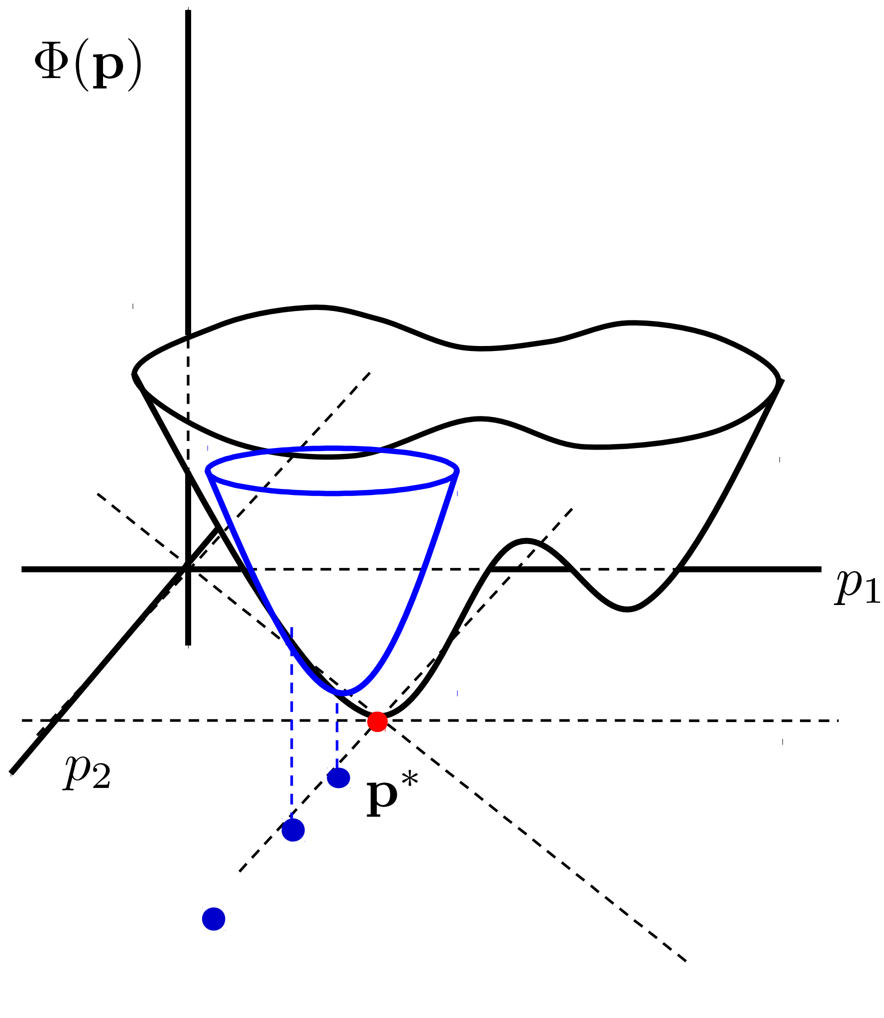


Approximate the nonlinear function around the initial approximation

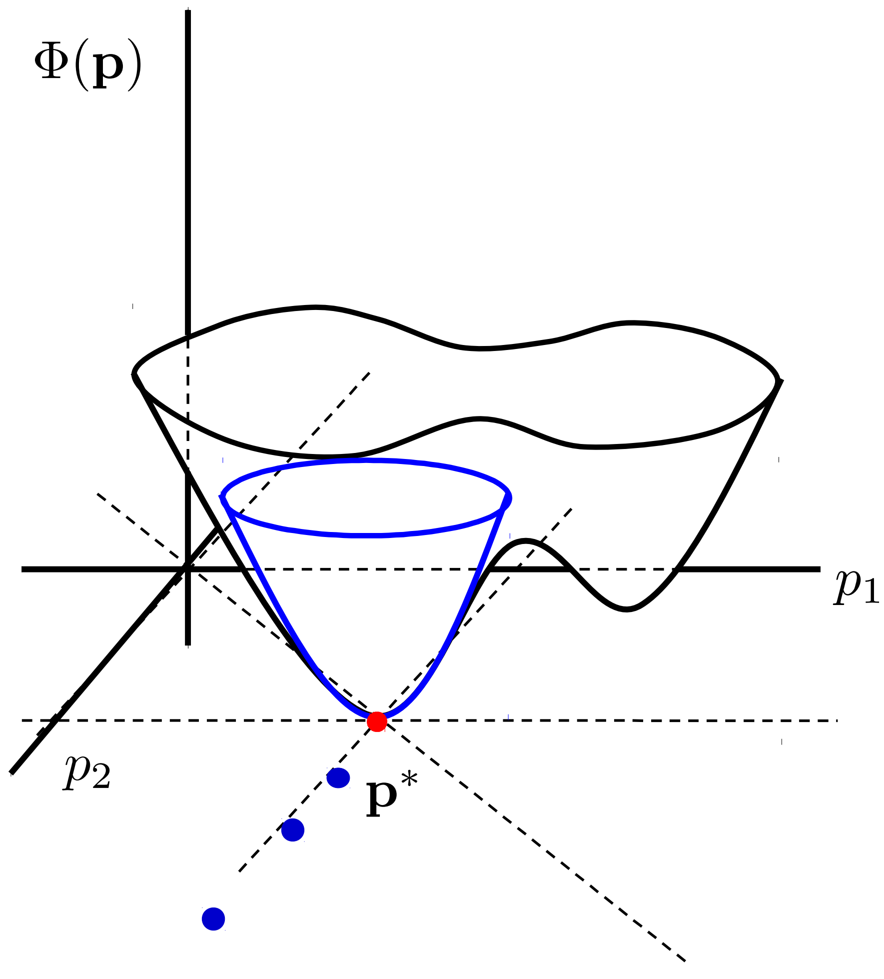


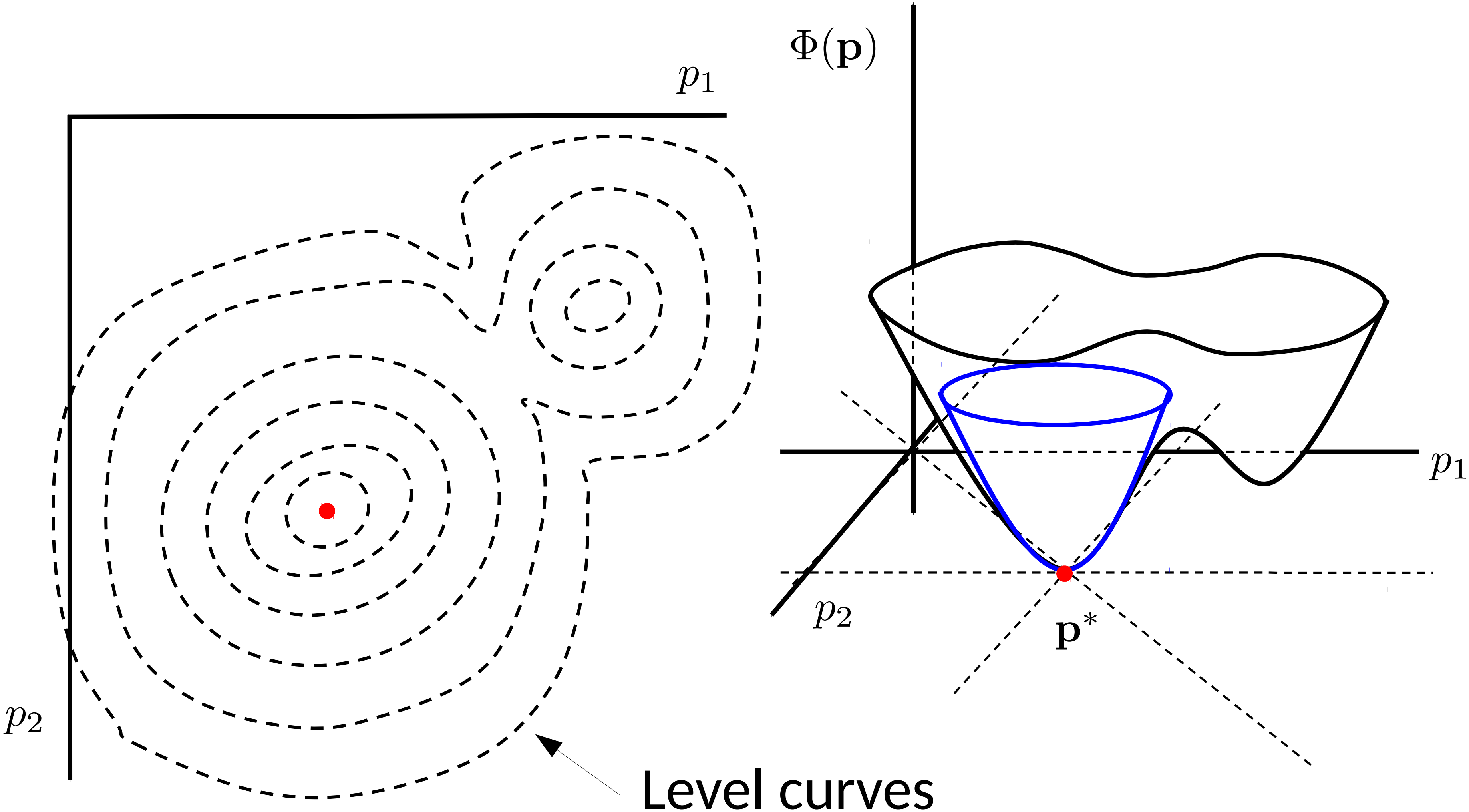
This approximation also has a minimum

Compute a new approximation around this minimum

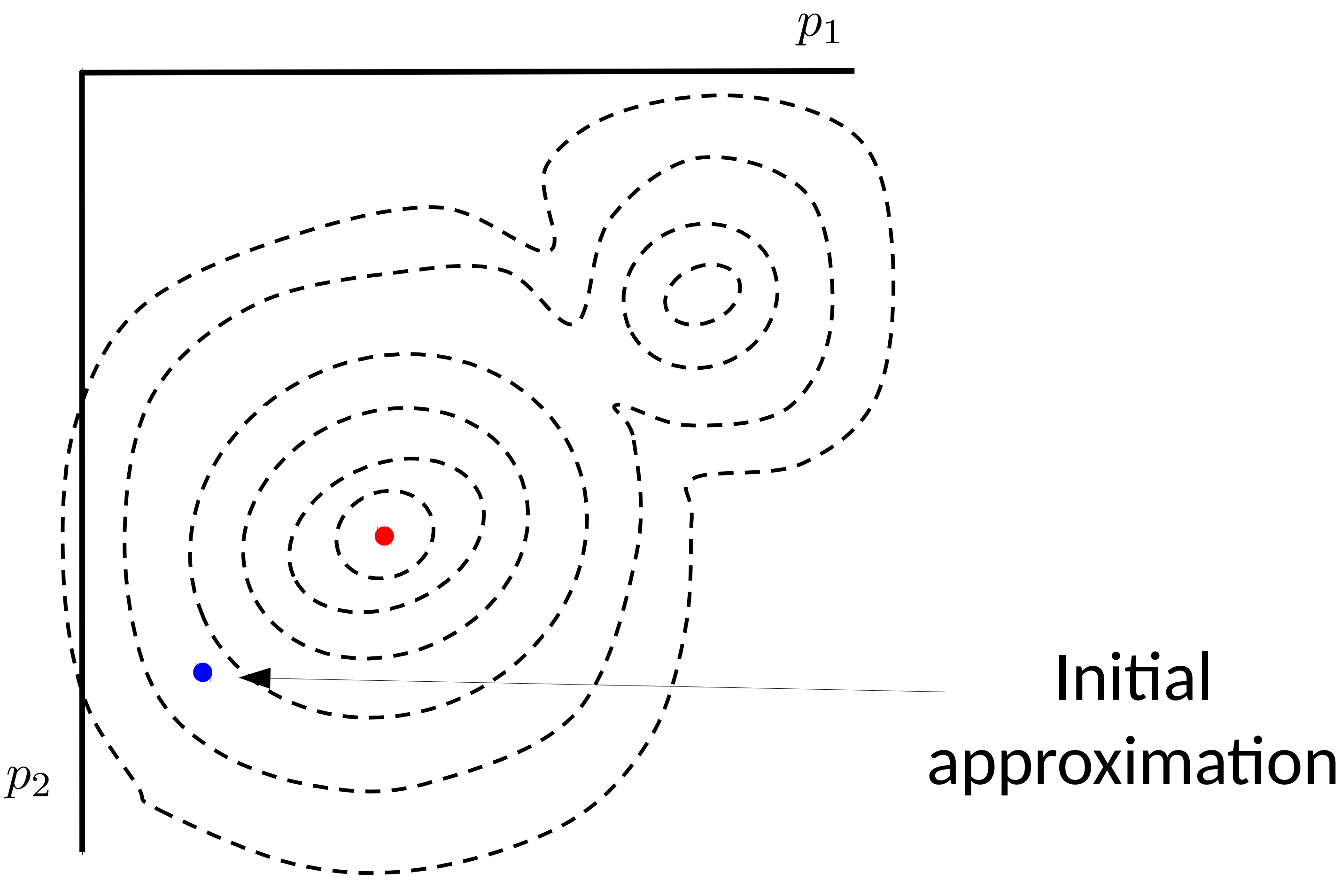


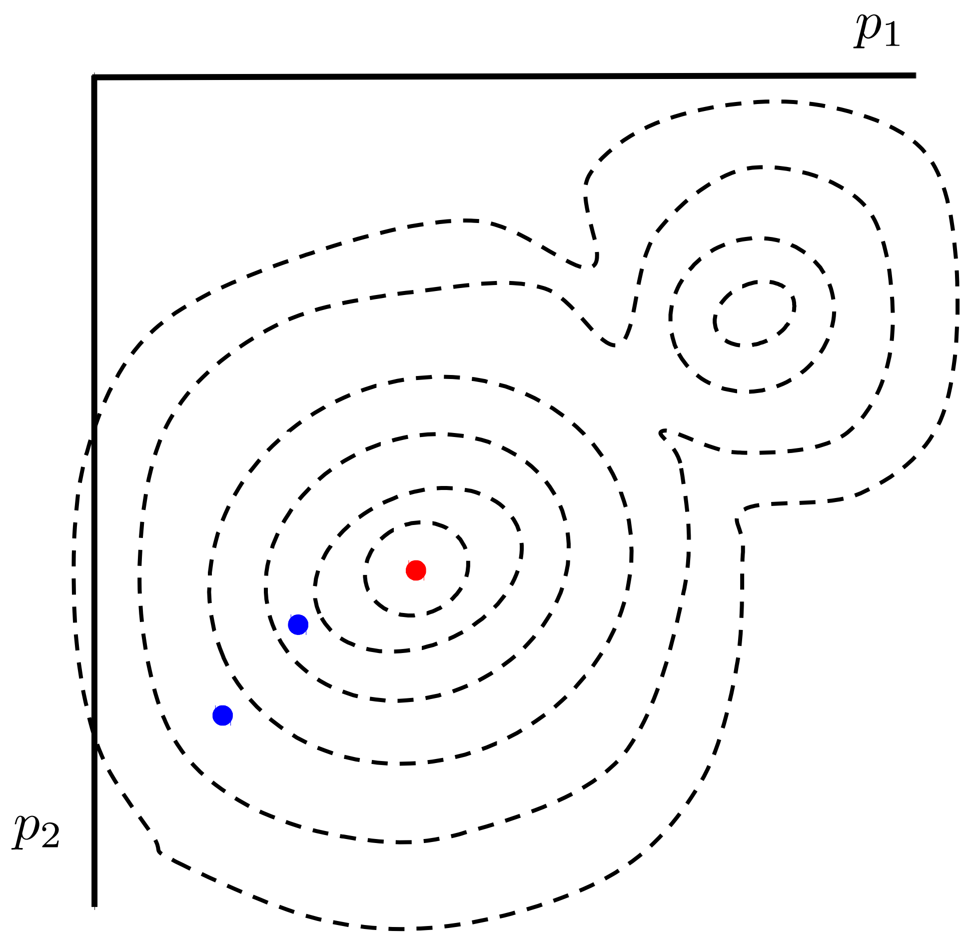
And so on ...



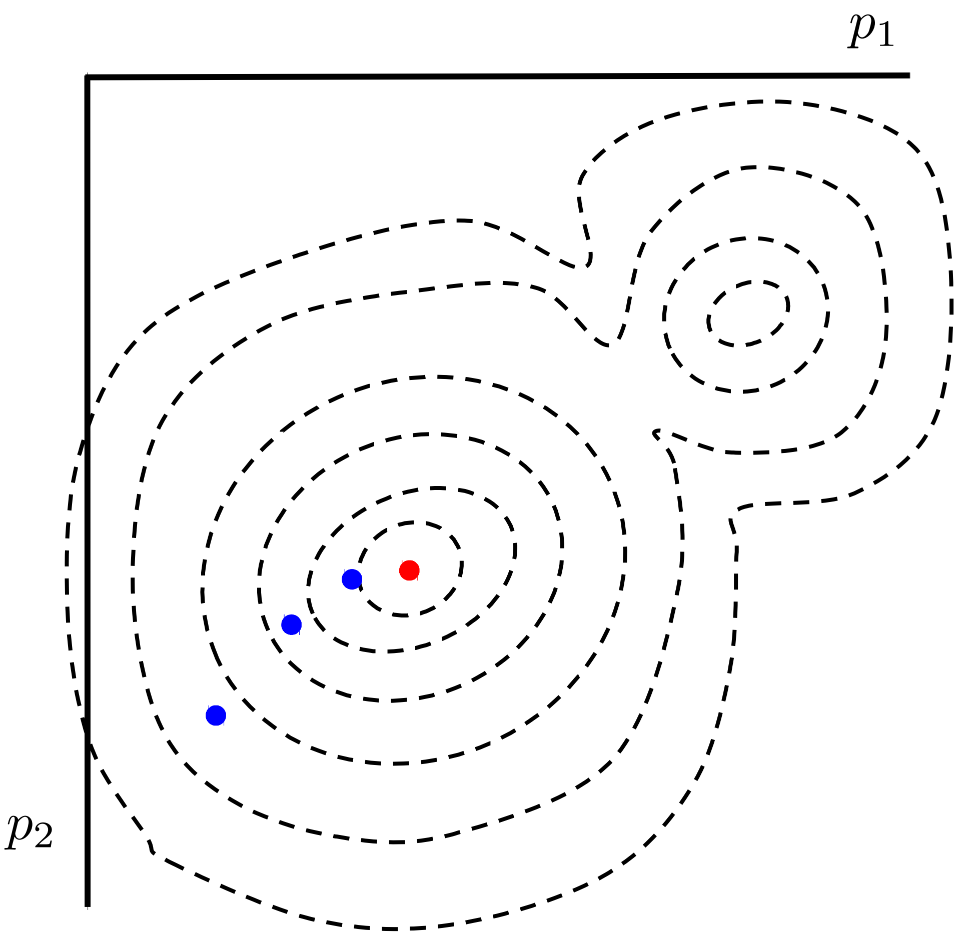


p_1
p_2
$\Phi(\mathbf{p})$
p_1
p_2
$\mathbf{p}^*$
Level curves

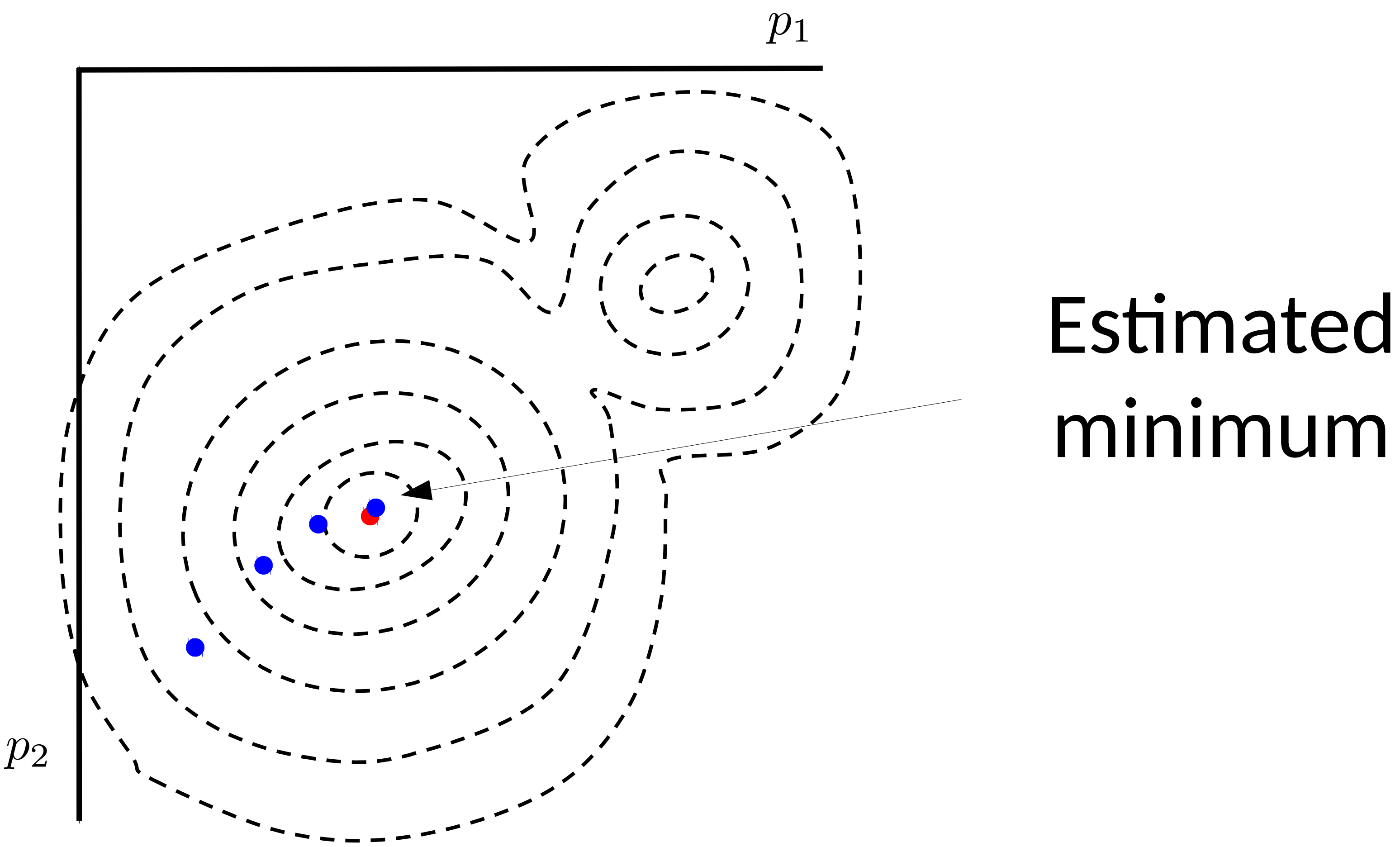

p_1
p_2
Initial approximation

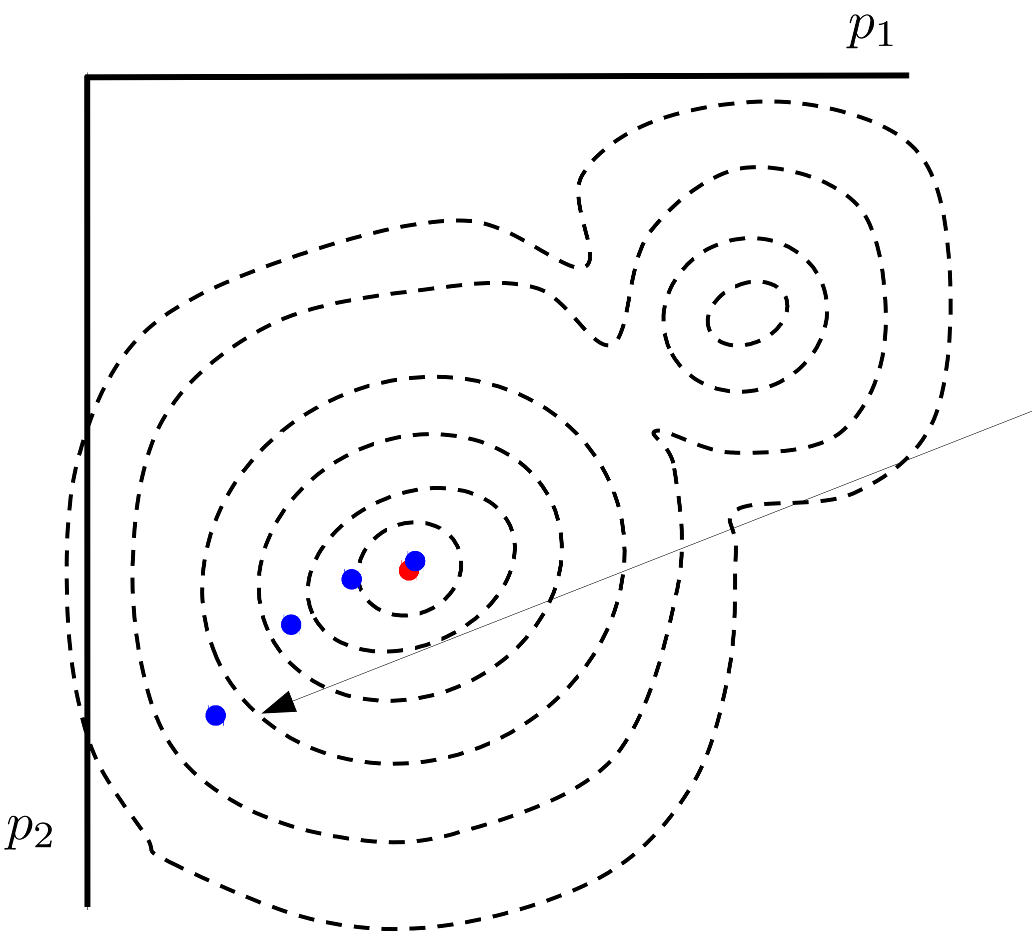


p_1
p_2

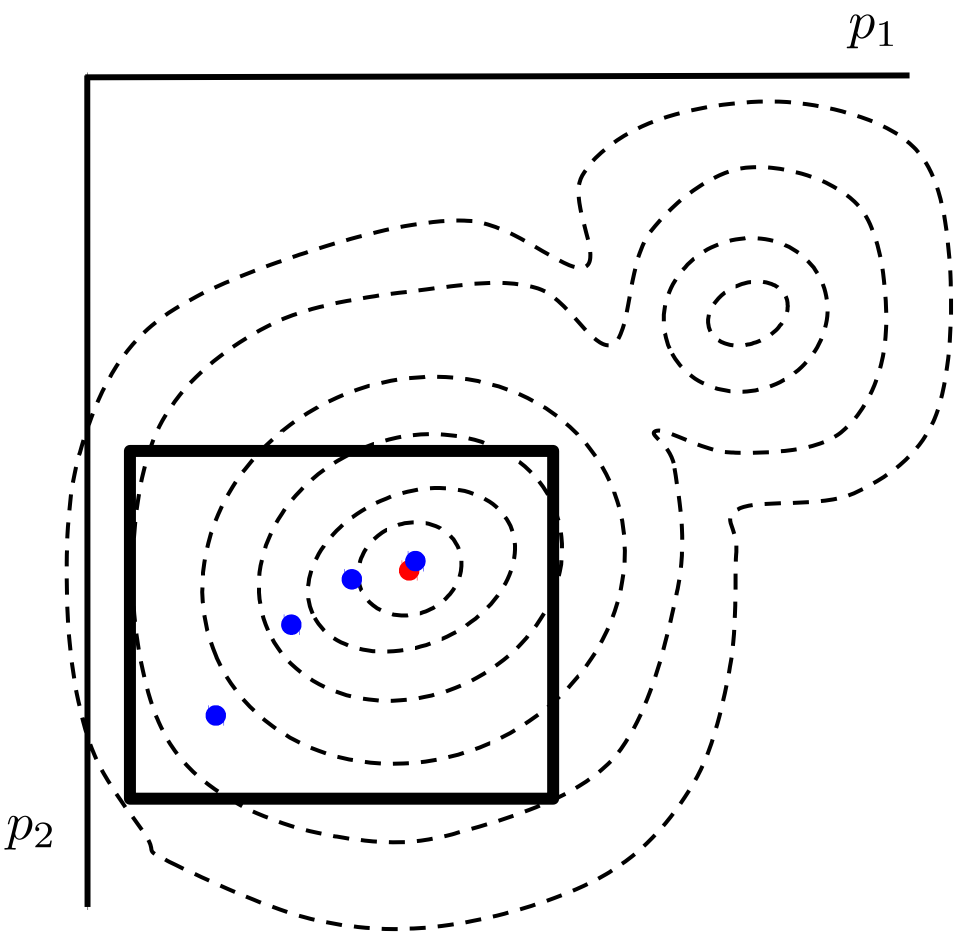


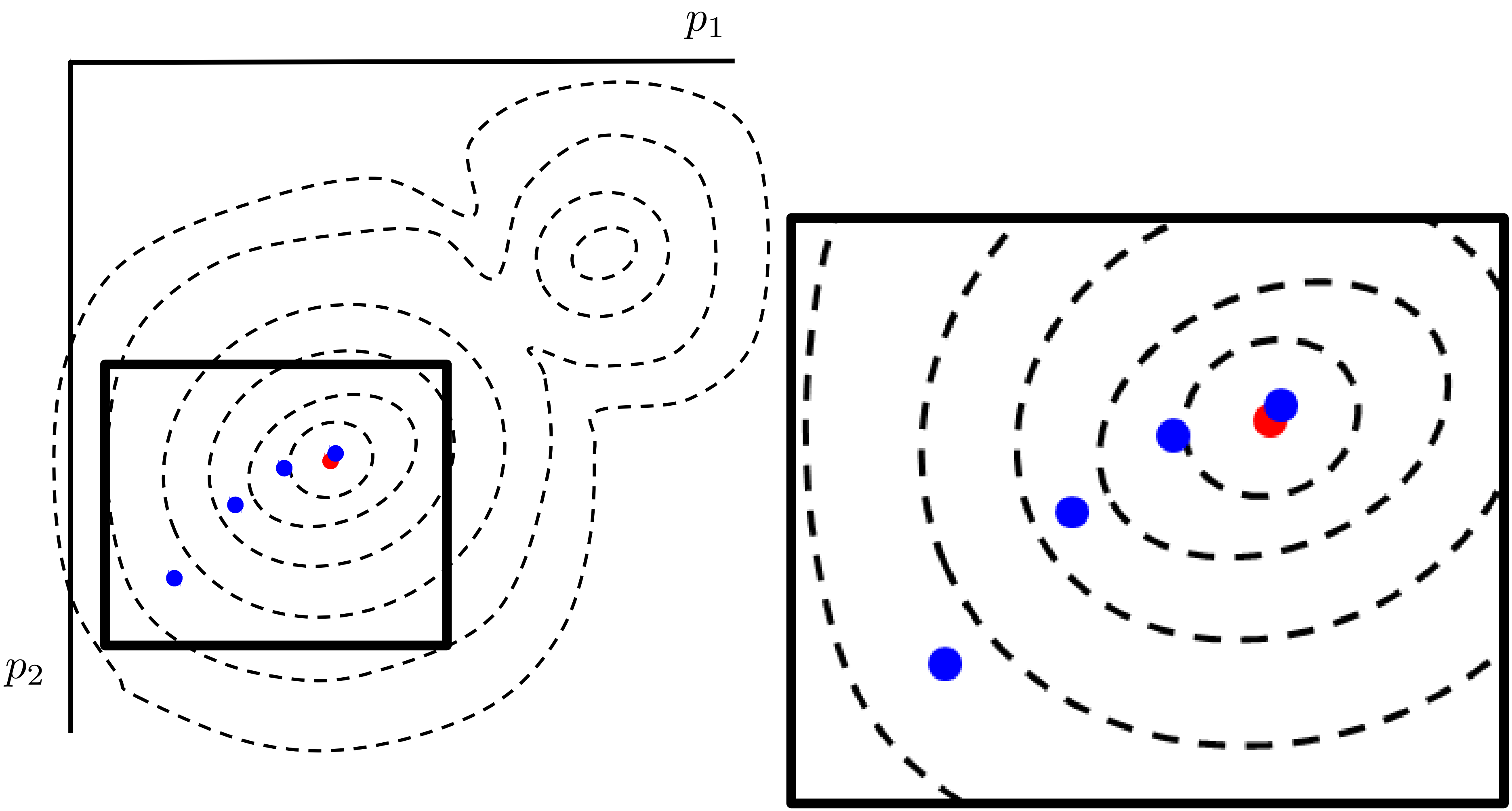
p_1
p_2

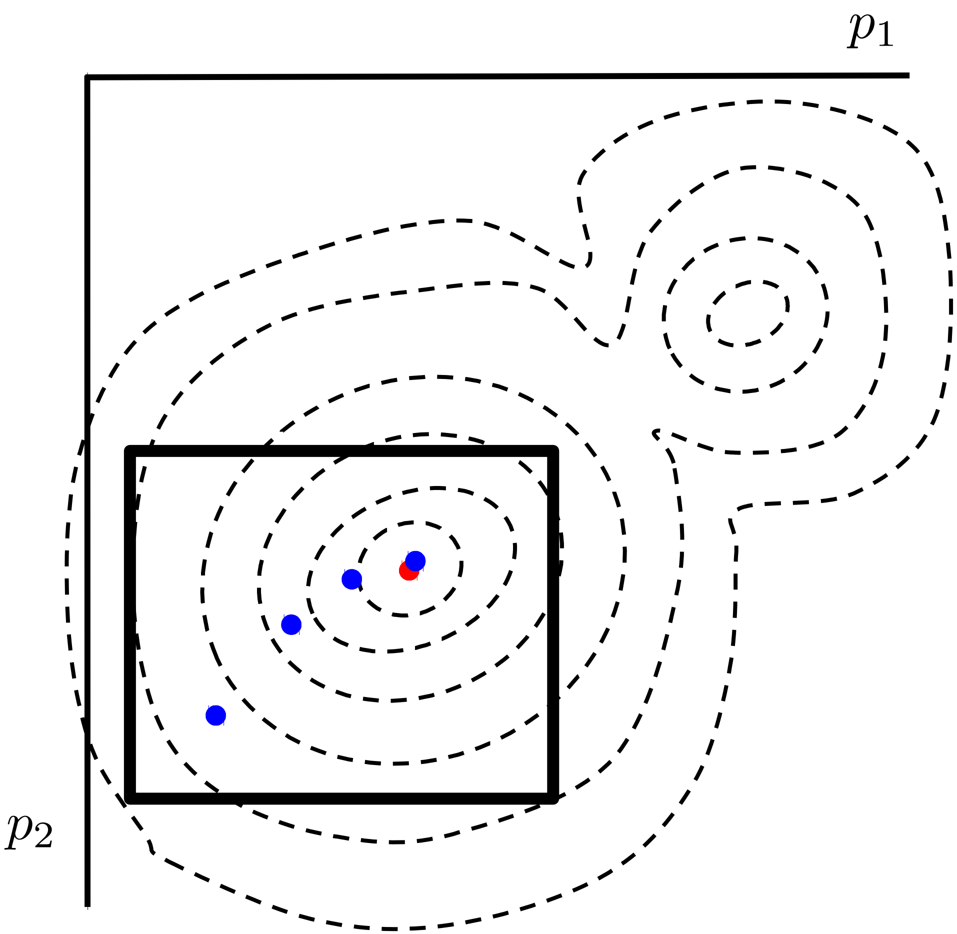

p_1
p_2
Estimated minimum



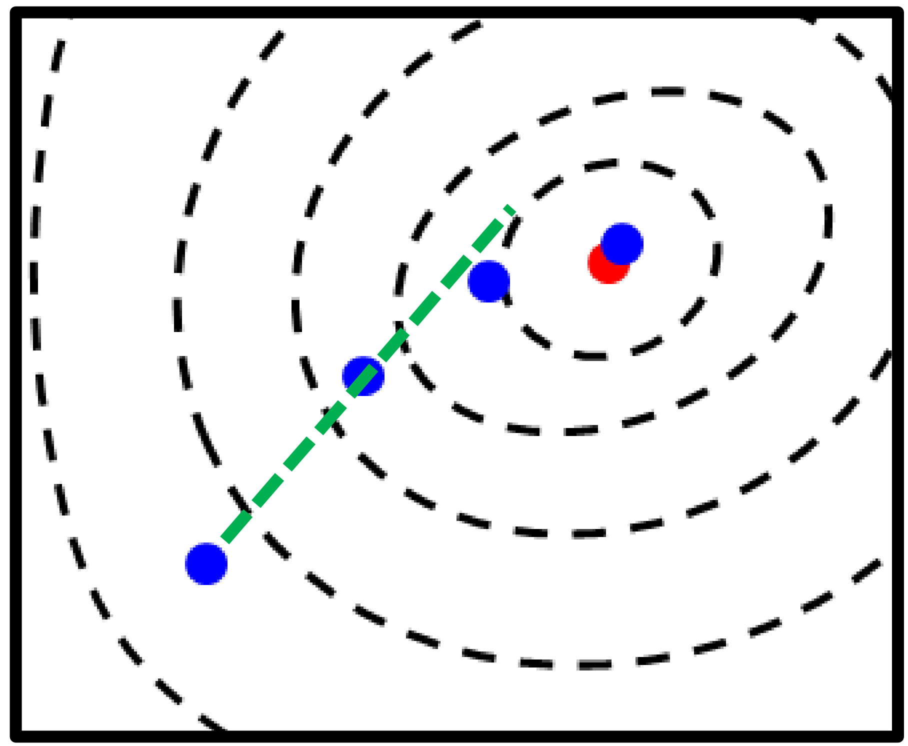
Given a point, it is needed to define a direction and the step length

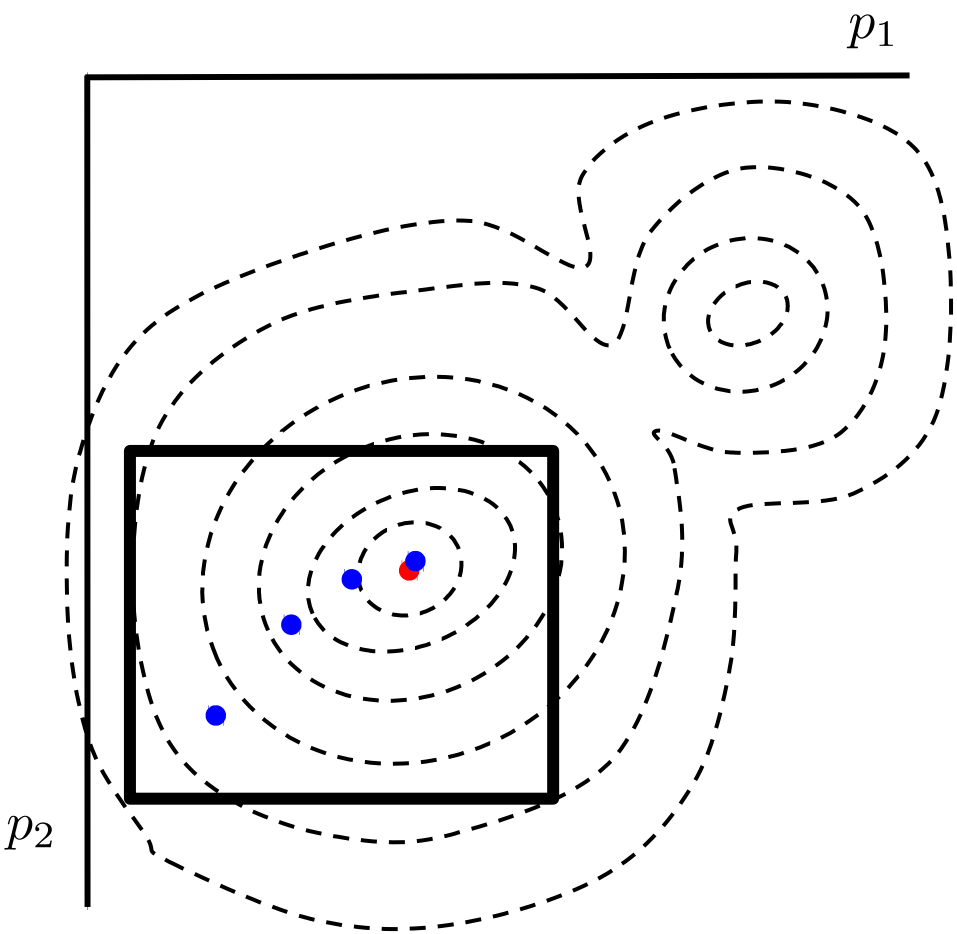

p_1
p_2


p_1
p_2

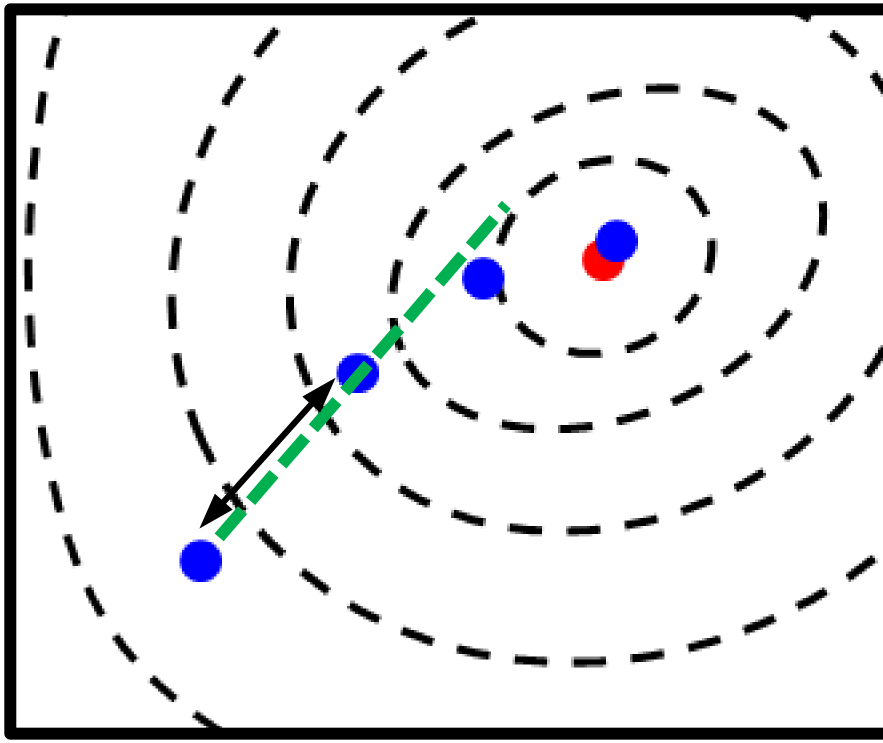


Direction

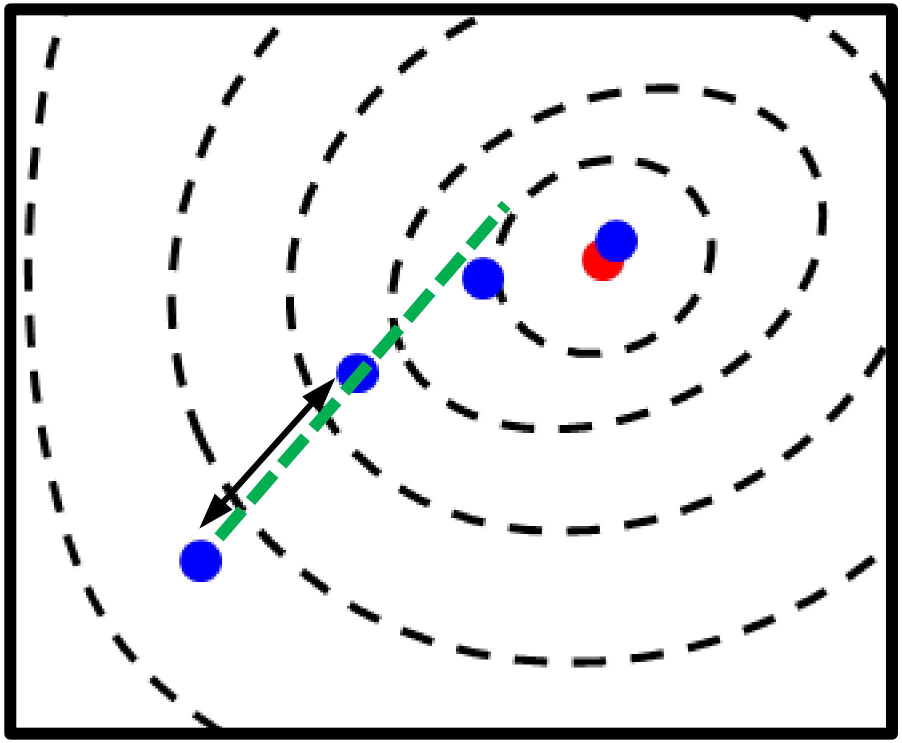




Step length



The direction and step length may be defined by using the gradient



$$\Phi(\mathbf{p})$$

$$\Phi(\mathbf{p})$$

$$\Phi(\mathbf{p}_0 + \Delta\mathbf{p}) \approx \Phi(\mathbf{p}_0) + \nabla\Phi(\mathbf{p}_0)^\top \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{p}^\top \mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p}$$

$$\Phi(\mathbf{p})$$

$$\Phi(\mathbf{p}_0 + \Delta\mathbf{p}) \approx \Phi(\mathbf{p}_0) + \nabla\Phi(\mathbf{p}_0)^\top \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{p}^\top \mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p} = -\nabla\Phi(\mathbf{p}_0)$$

$$\Phi(\mathbf{p})$$

$$\Phi(\mathbf{p}_0 + \Delta\mathbf{p}) \approx \Phi(\mathbf{p}_0) + \nabla\Phi(\mathbf{p}_0)^\top \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{p}^\top \boxed{\mathbf{H}(\mathbf{p}_0)} \Delta\mathbf{p}$$

$$\boxed{\mathbf{H}(\mathbf{p}_0)} \Delta\mathbf{p} = -\nabla\Phi(\mathbf{p}_0)$$

Difference between the methods

$$\Phi(\mathbf{p})$$

$$\Phi(\mathbf{p}_0 + \Delta\mathbf{p}) \approx \Phi(\mathbf{p}_0) + \nabla\Phi(\mathbf{p}_0)^\top \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{p}^\top \boxed{\mathbf{H}(\mathbf{p}_0)} \Delta\mathbf{p}$$

$$\boxed{\mathbf{H}(\mathbf{p}_0)} \Delta\mathbf{p} = -\nabla\Phi(\mathbf{p}_0)$$

Newton	$\mathbf{H}(\mathbf{p}_0)$
Gauss - Newton	$\tilde{\mathbf{H}}(\mathbf{p}_0)$
Steepest decent	$\lambda\mathbf{I}$
Levenberg - Marquardt	$\tilde{\mathbf{H}}(\mathbf{p}_0) + \lambda\mathbf{I}$

Difference between the methods

Method	Convergence
Steepest Decent	0
Levenberg - Marquardt	1
Gauss - Newton	2
Newton	3

0 – slow 3 – fast

Method	Initial approx
Steepest Decent	Can be distant
Levenberg - Marquardt	Can be distant
Gauss - Newton	Must be close
Newton	Must be close

Method	Direction/ Step length
Steepest Decent	Defined by the gradient
Levenberg - Marquardt	Defined by the Hessian and gradient
Gauss - Newton	Defined by the Hessian and gradient
Newton	Defined by the Hessian and gradient

Method	Computational cost
Steepest Decent	0
Levenberg - Marquardt	2
Gauss - Newton	1
Newton	3

0 – low 3 – high