

Topics on nonlinear optimization

Contents

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

Introduction

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1} \qquad \bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1} \qquad \bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

$$\bar{p}=\begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M\times 1}$$

$$\overline{\nabla}\phi(\bar{p}^*)=\overline{0}_{M\times 1}$$

$$\bar{d}=\begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N\times 1} \qquad \bar{g}(\bar{p})=\begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N\times 1}$$

$$\phi(\bar{p})=[\bar{d}-\bar{g}(\bar{p})]^T[\bar{d}-\bar{g}(\bar{p})]$$

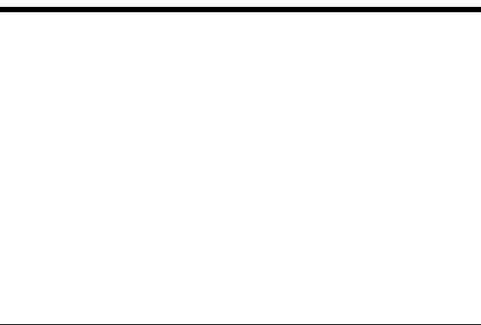
$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1} \quad \bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

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

Linear problem



Nonlinear problem



$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

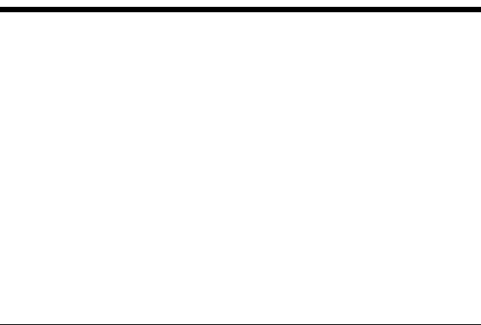
$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

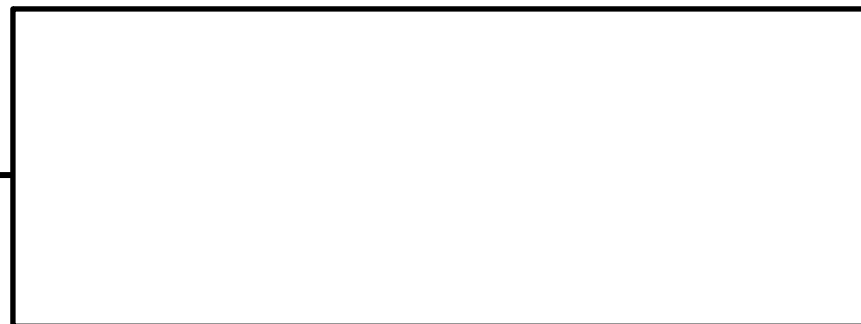
$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

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

Linear problem



Nonlinear problem



$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

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

Linear problem

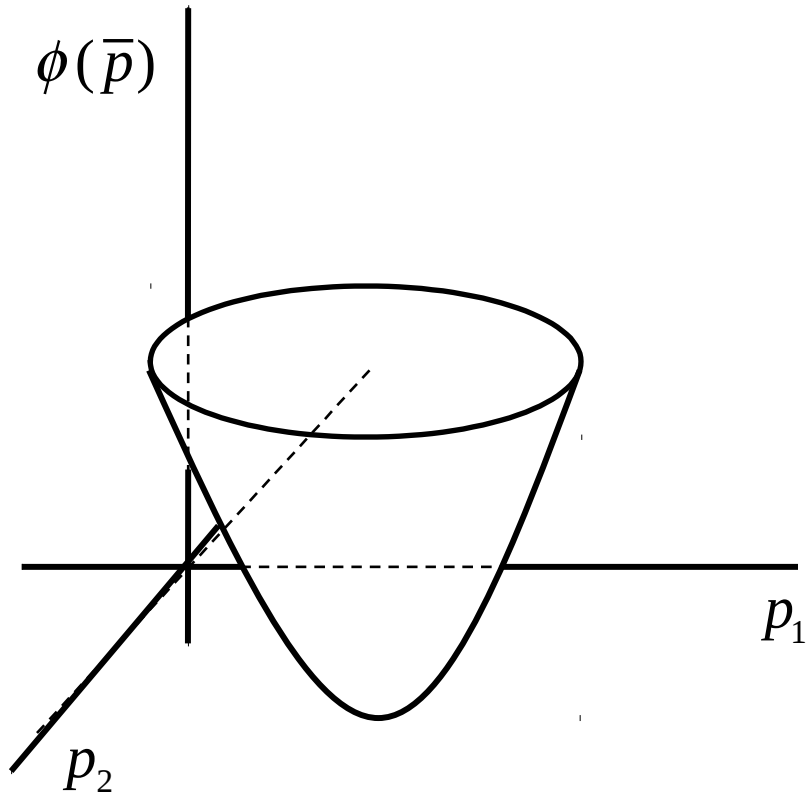
$$\bar{g}(\bar{p}) = \bar{\bar{B}}\bar{p} + \bar{b}$$

$$\bar{p}^* = \left(\begin{bmatrix} \bar{\bar{B}} & \bar{\bar{B}} \end{bmatrix}^T \right)^{-1} \bar{\bar{B}}^T [\bar{d} - \bar{b}]$$

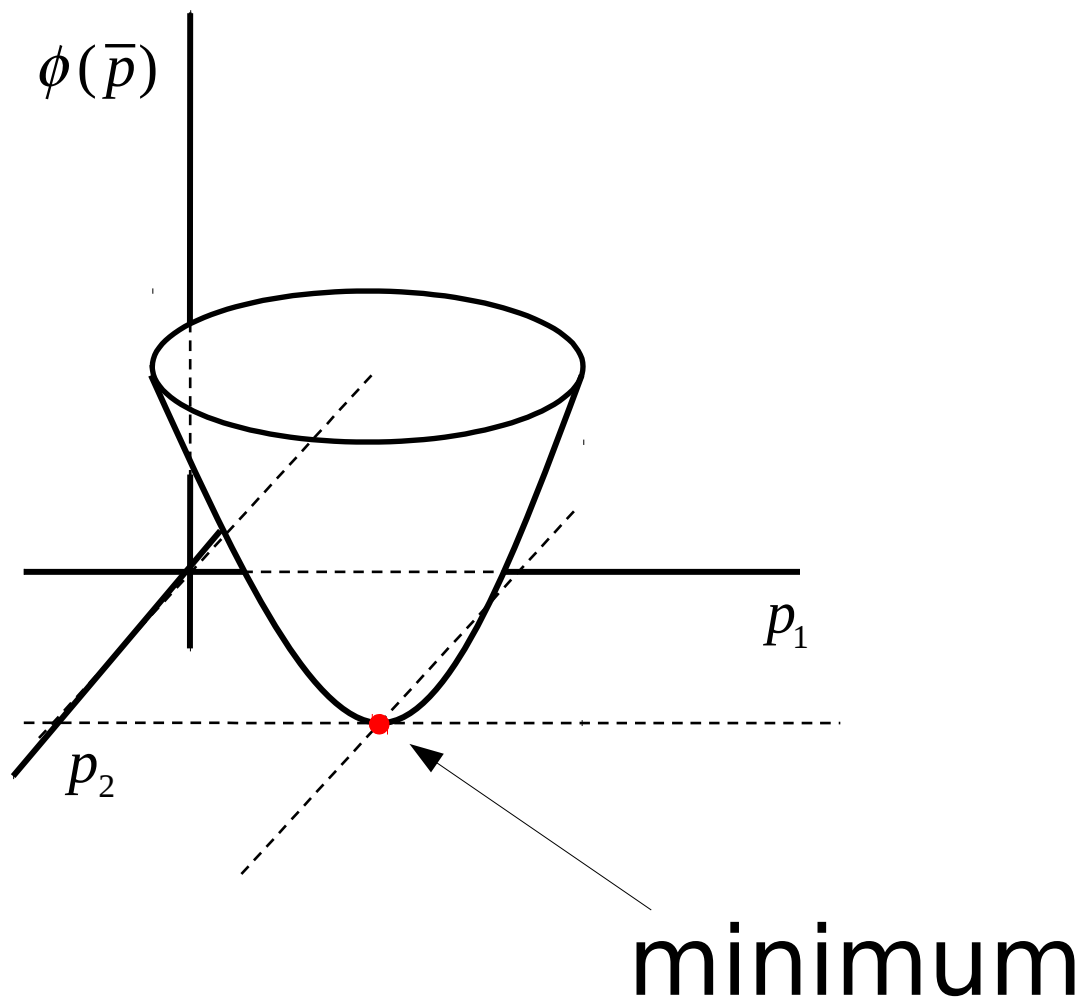
Nonlinear problem

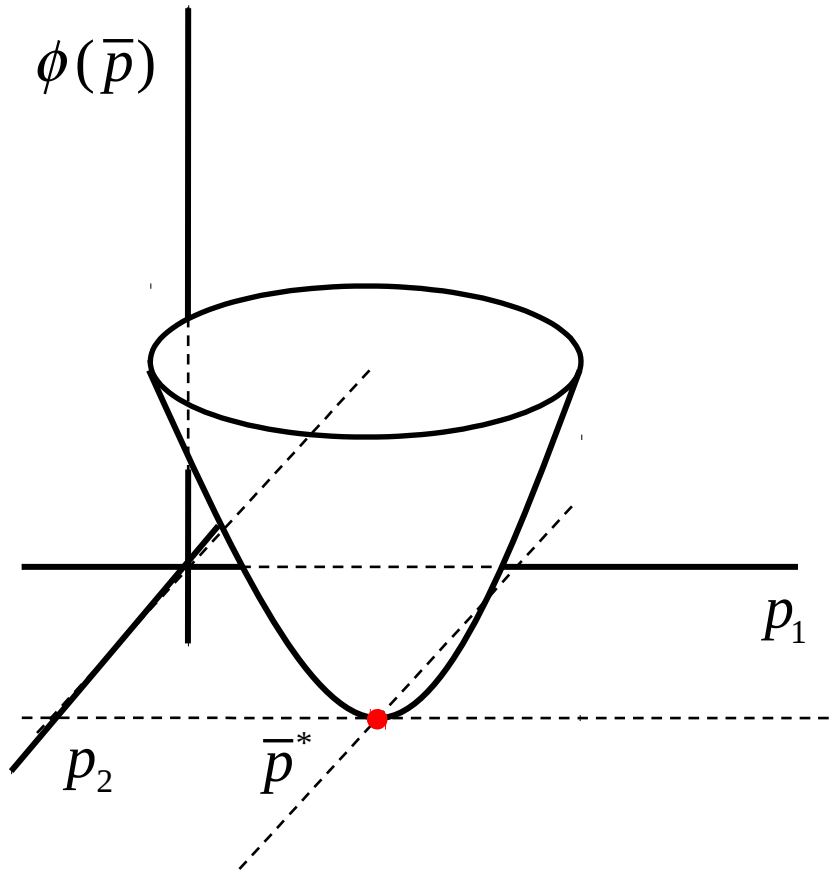
$$\bar{g}(\bar{p}) \neq \bar{\bar{B}}\bar{p} + \bar{b}$$

$$\Delta \bar{p} = \left(\bar{\bar{G}}(\bar{p}_0)^T \bar{\bar{G}}(\bar{p}_0) \right)^{-1} \bar{\bar{G}}(\bar{p}_0)^T [\bar{d} - \bar{g}(\bar{p}_0)]$$

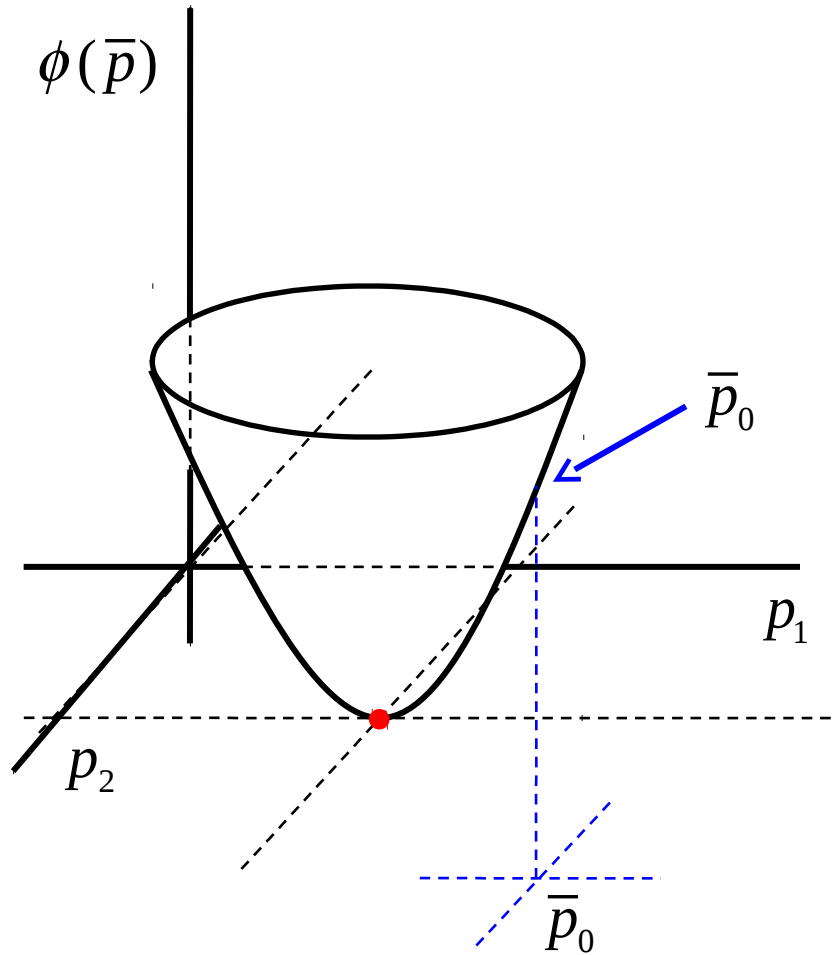


Linear
problem

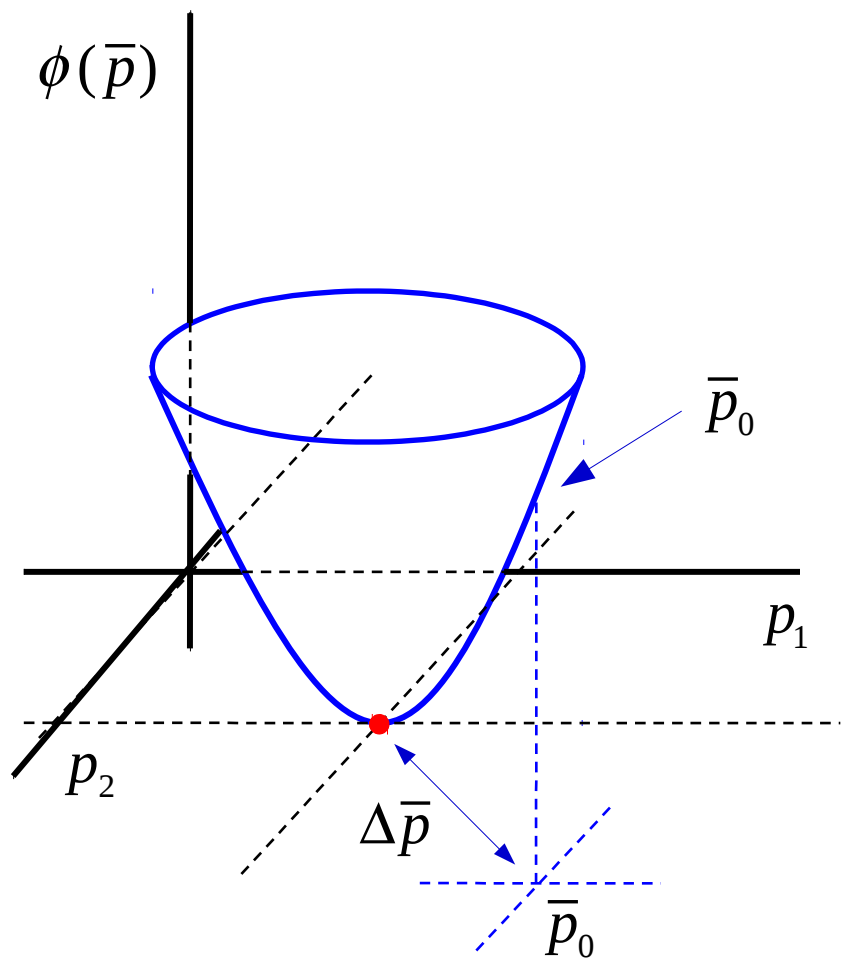


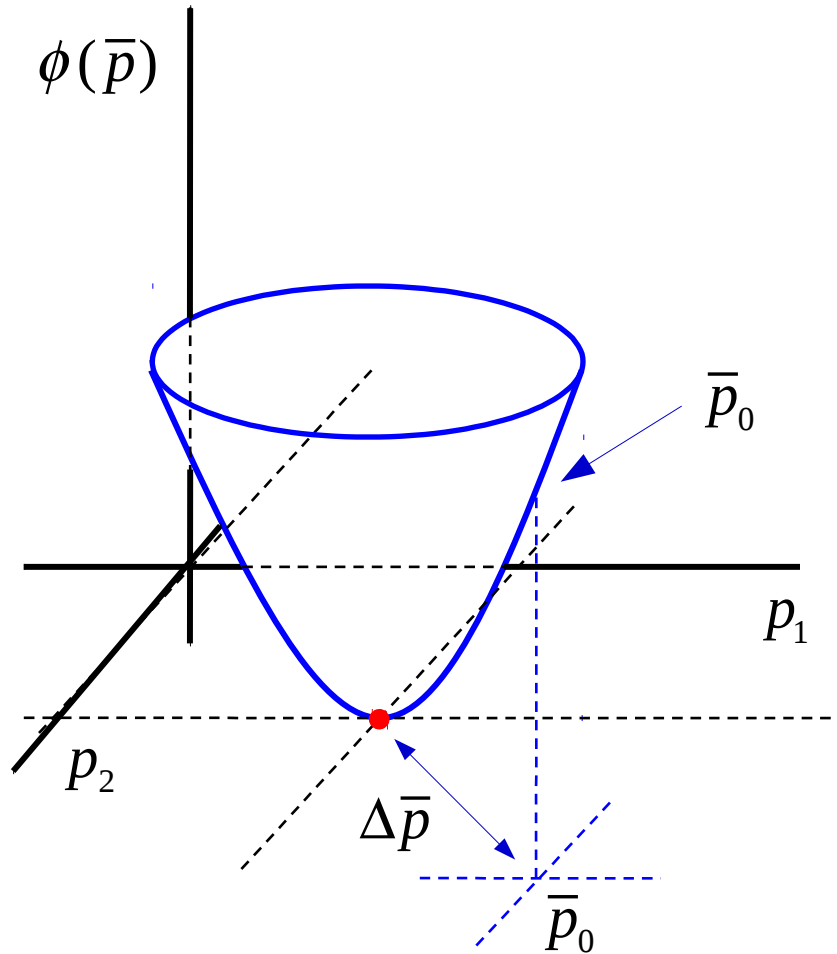


The minimum
can be
computed in
a single step



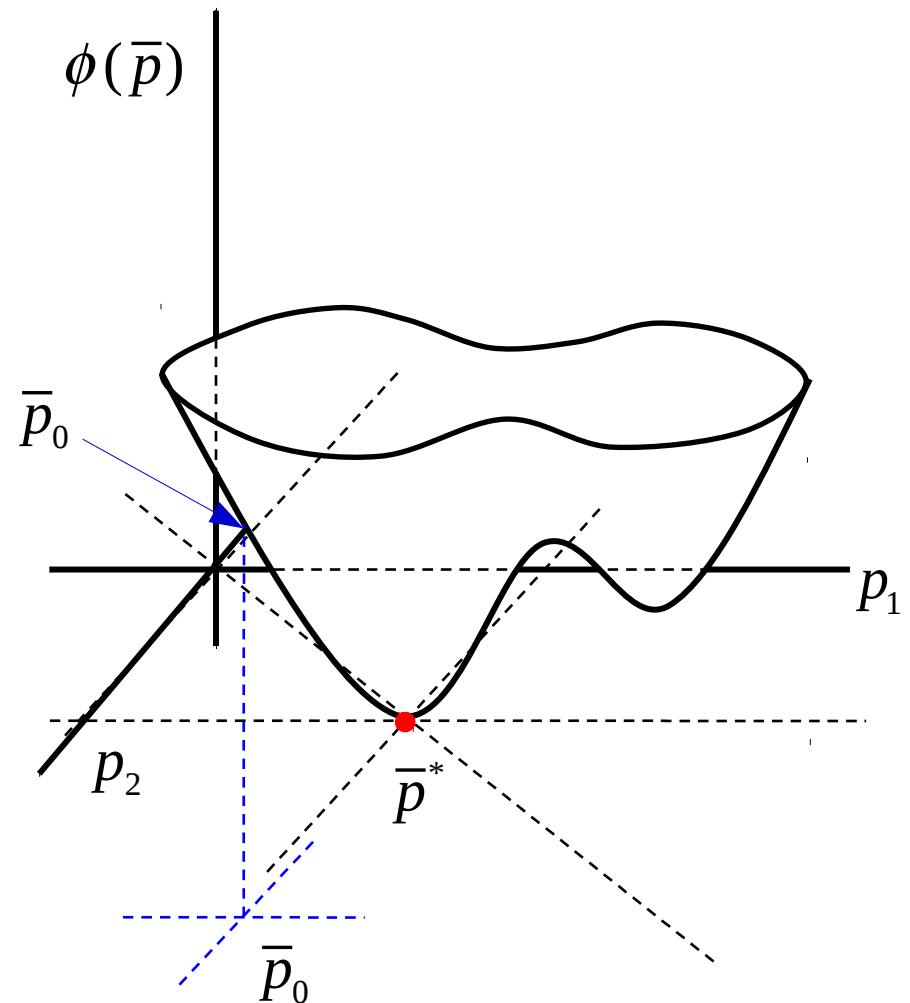
Or iteratively,
from a given
initial
approximation



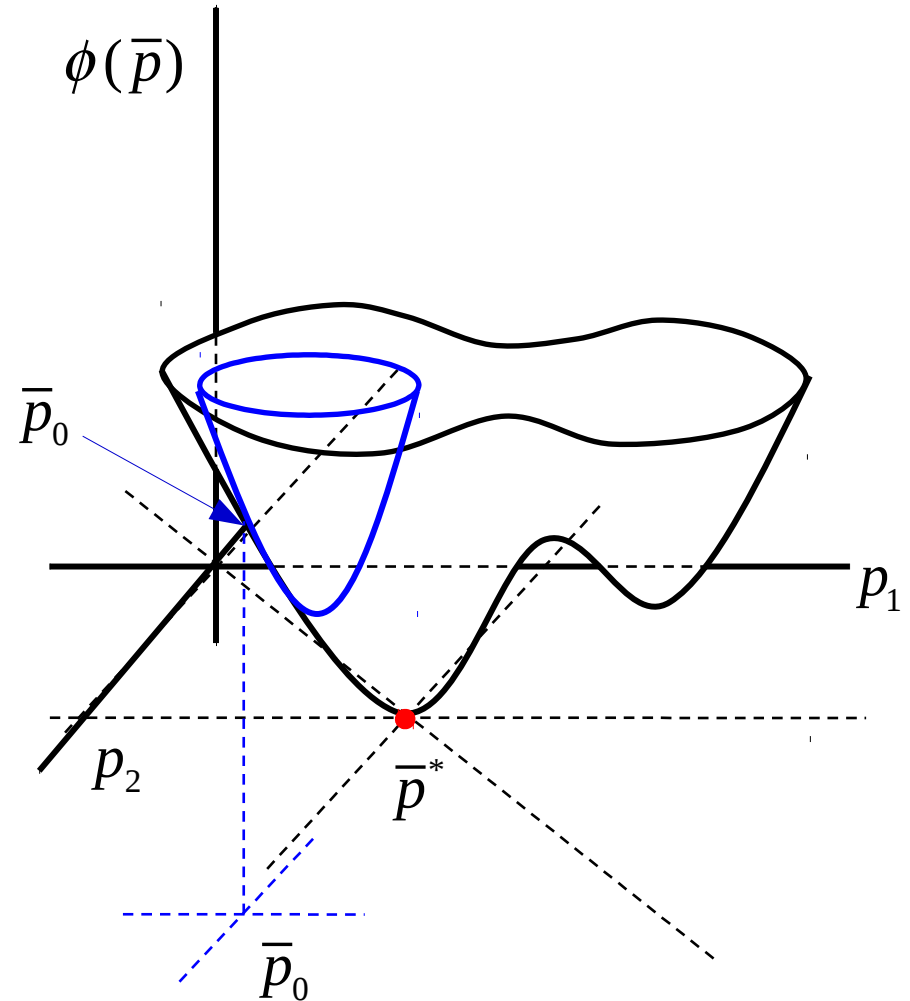


In this case,
the minimum
is estimated in
a single step
from the given
initial
approximation

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

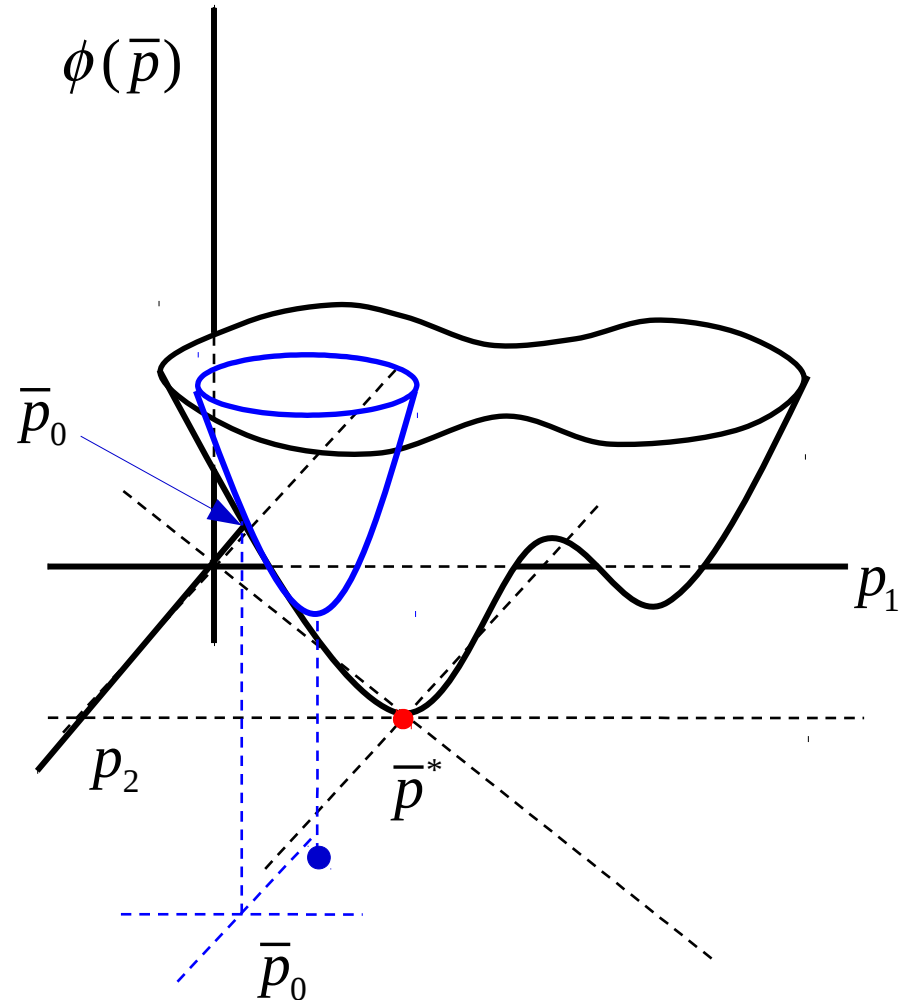


Approximate the
nonlinear
function around
the initial
approximation

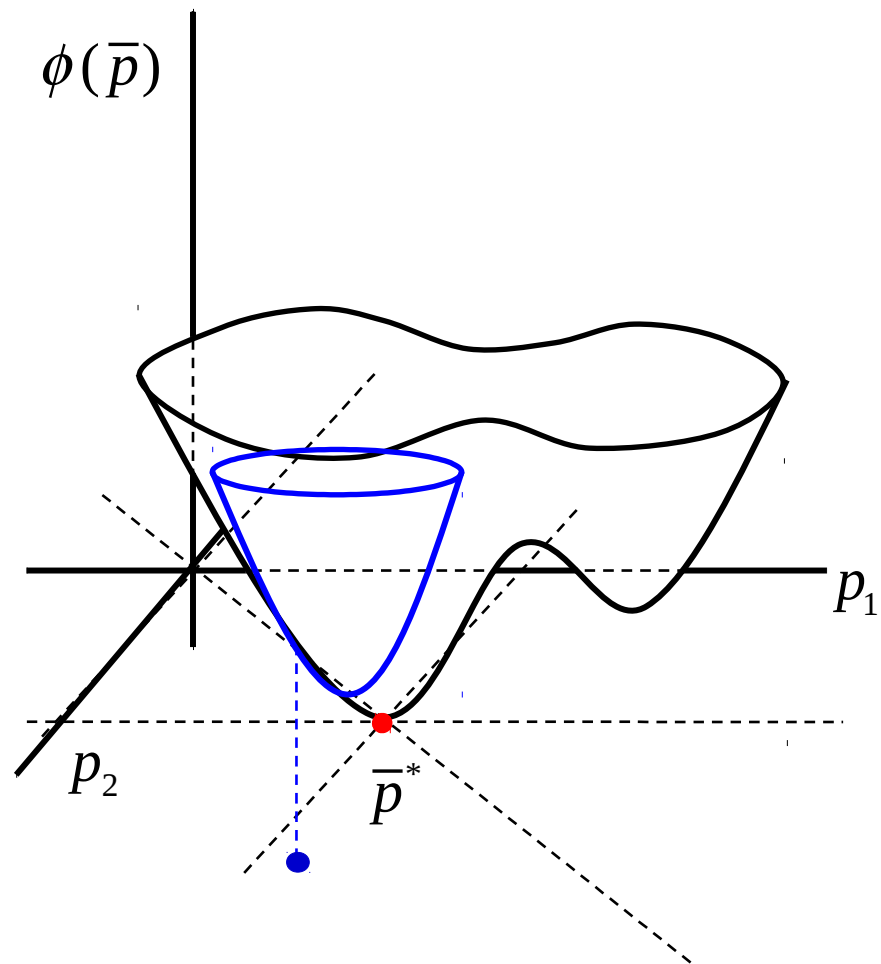


Approximate the
nonlinear
function around
the initial
approximation

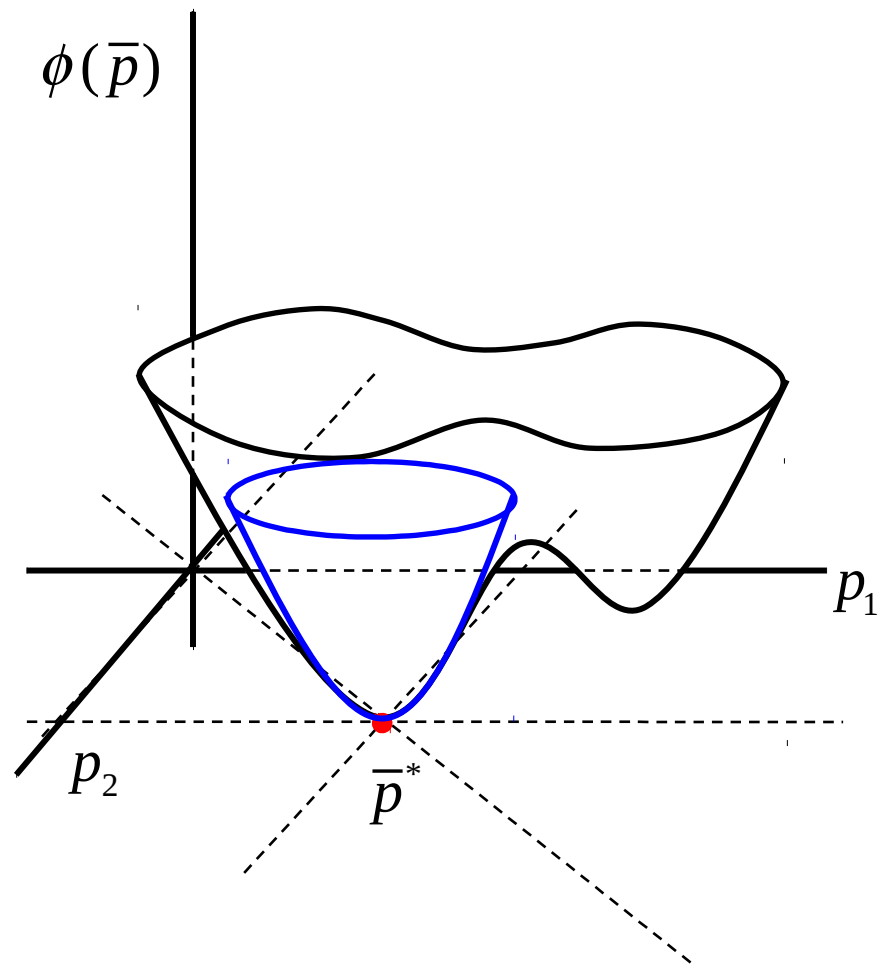
This
approximation also
has a minimum

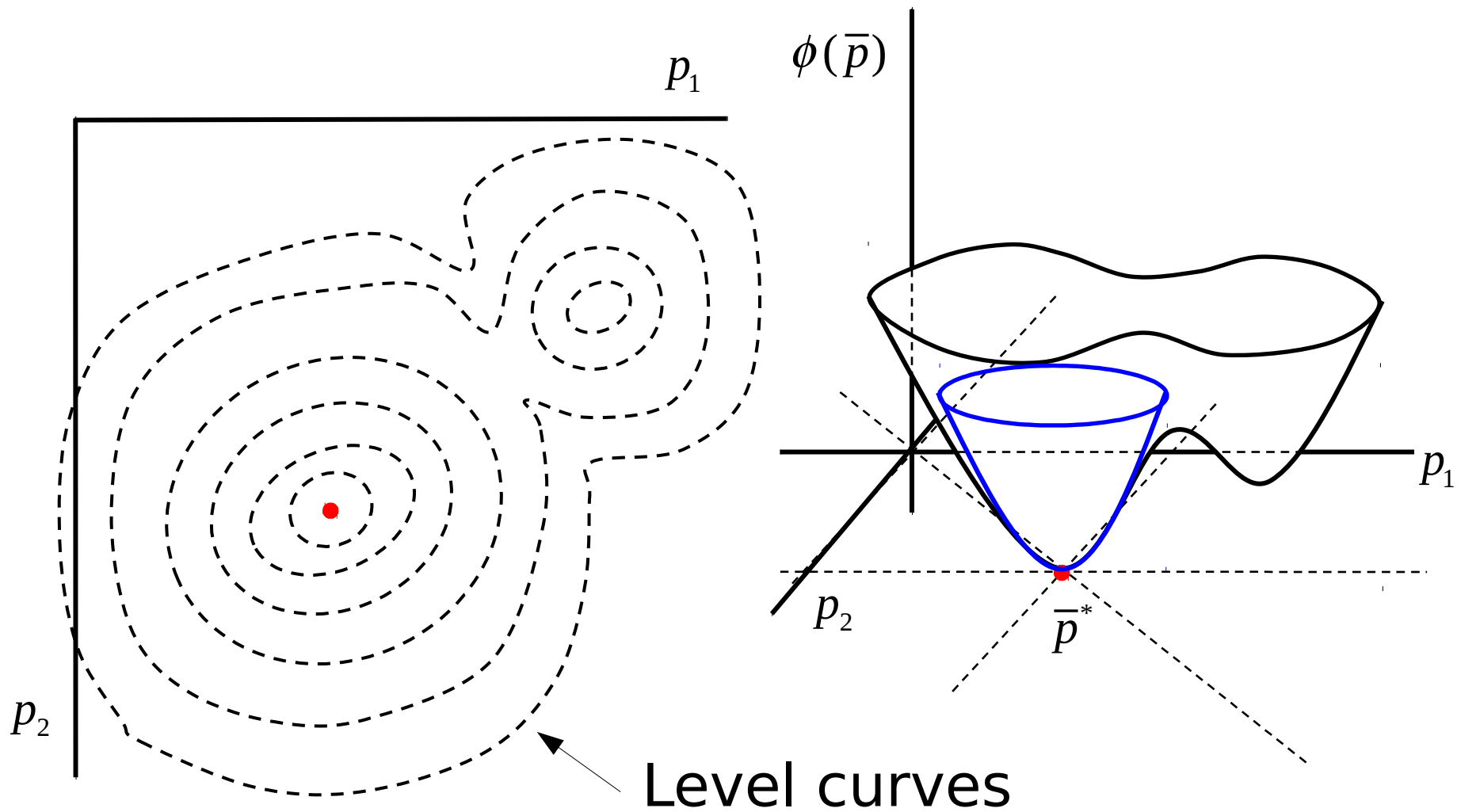


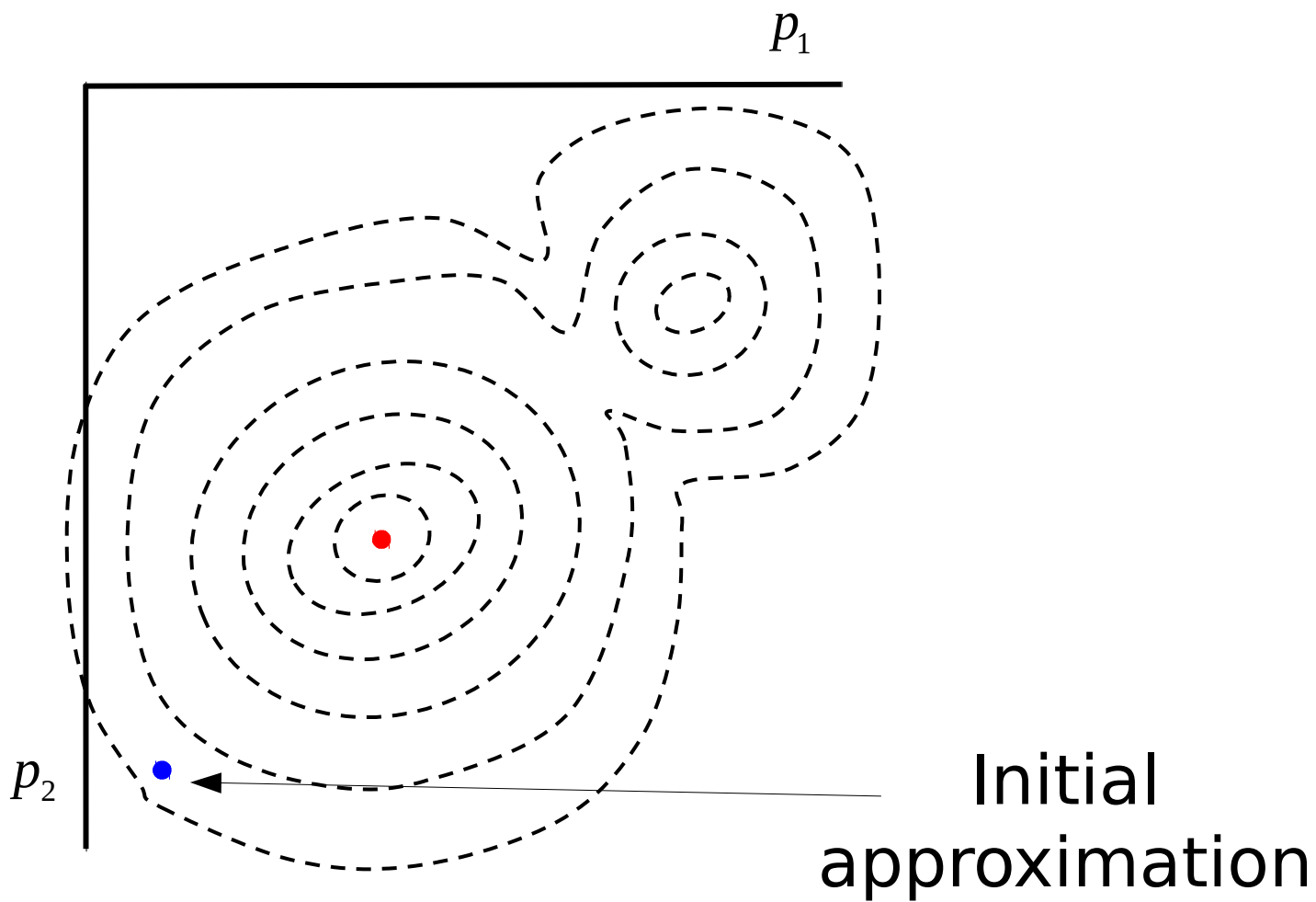
Compute a new
approximation
around this
minimum

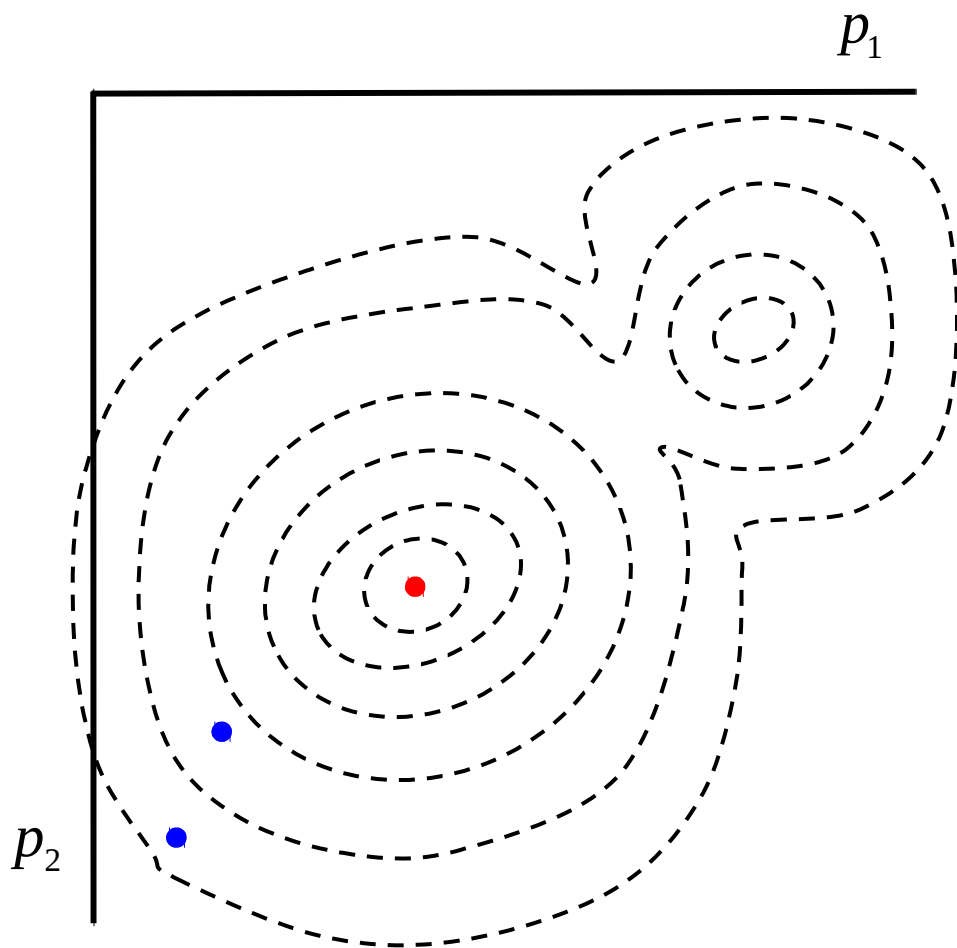


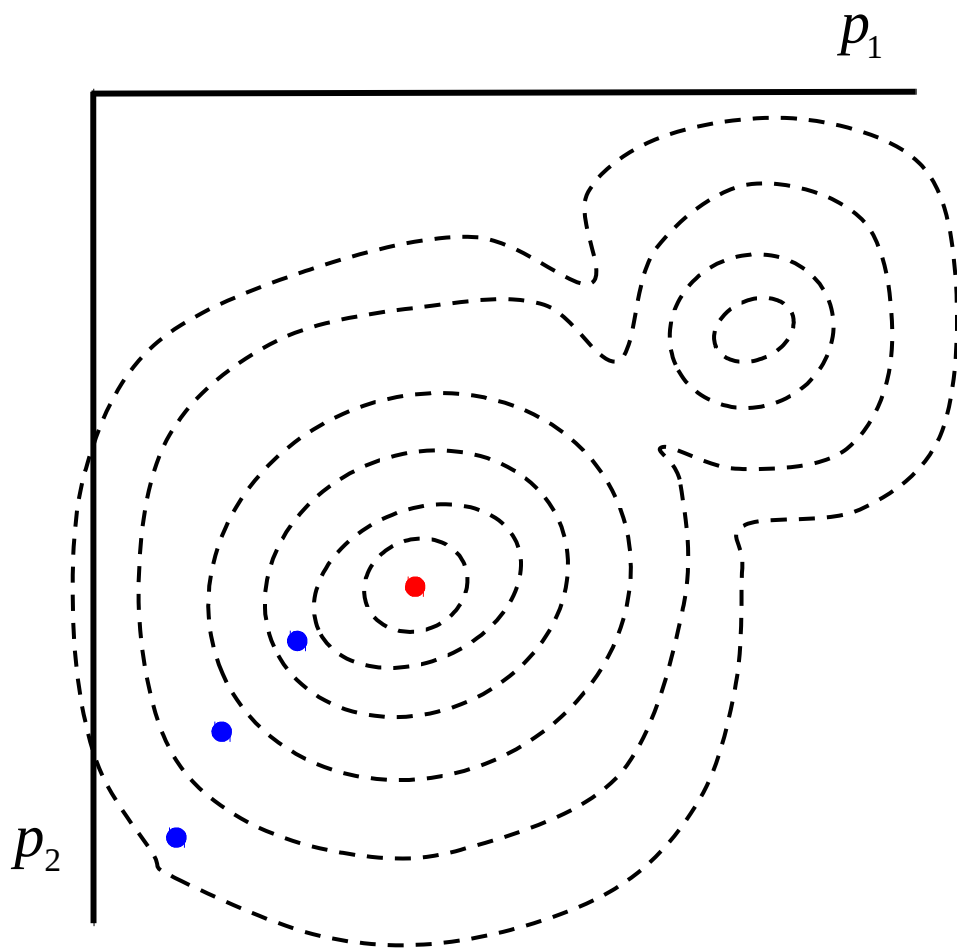
And so on ...

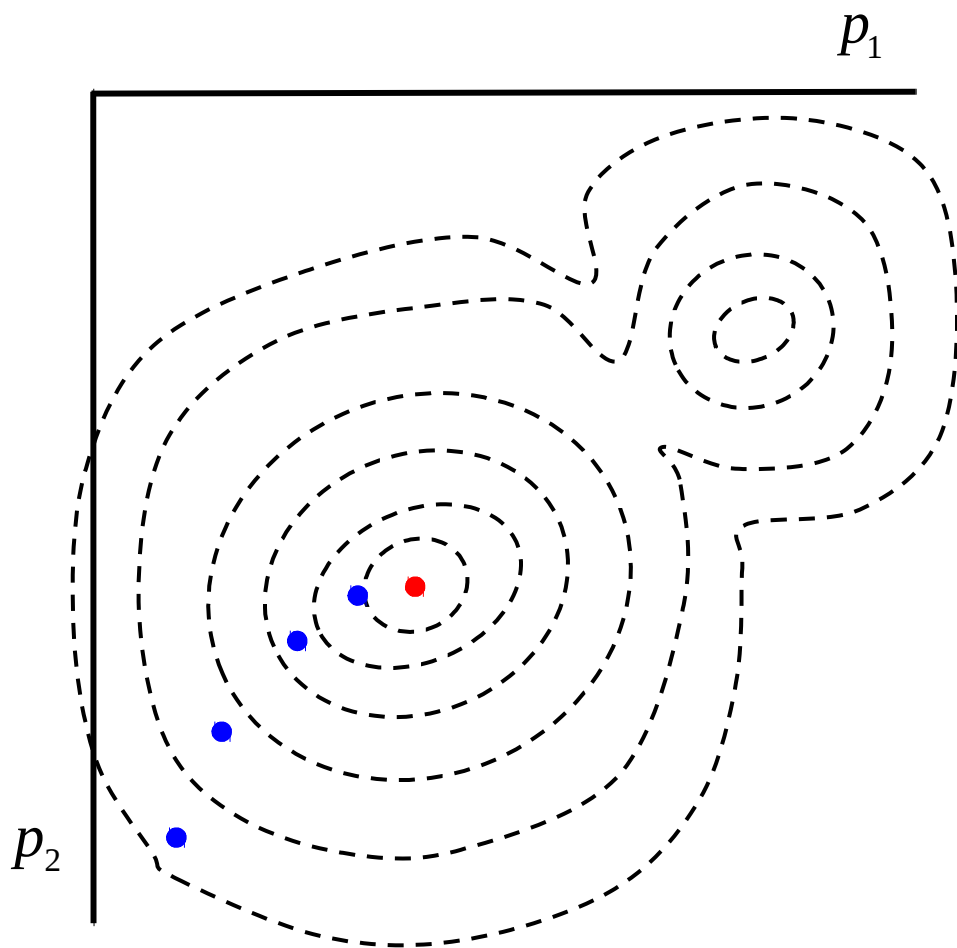


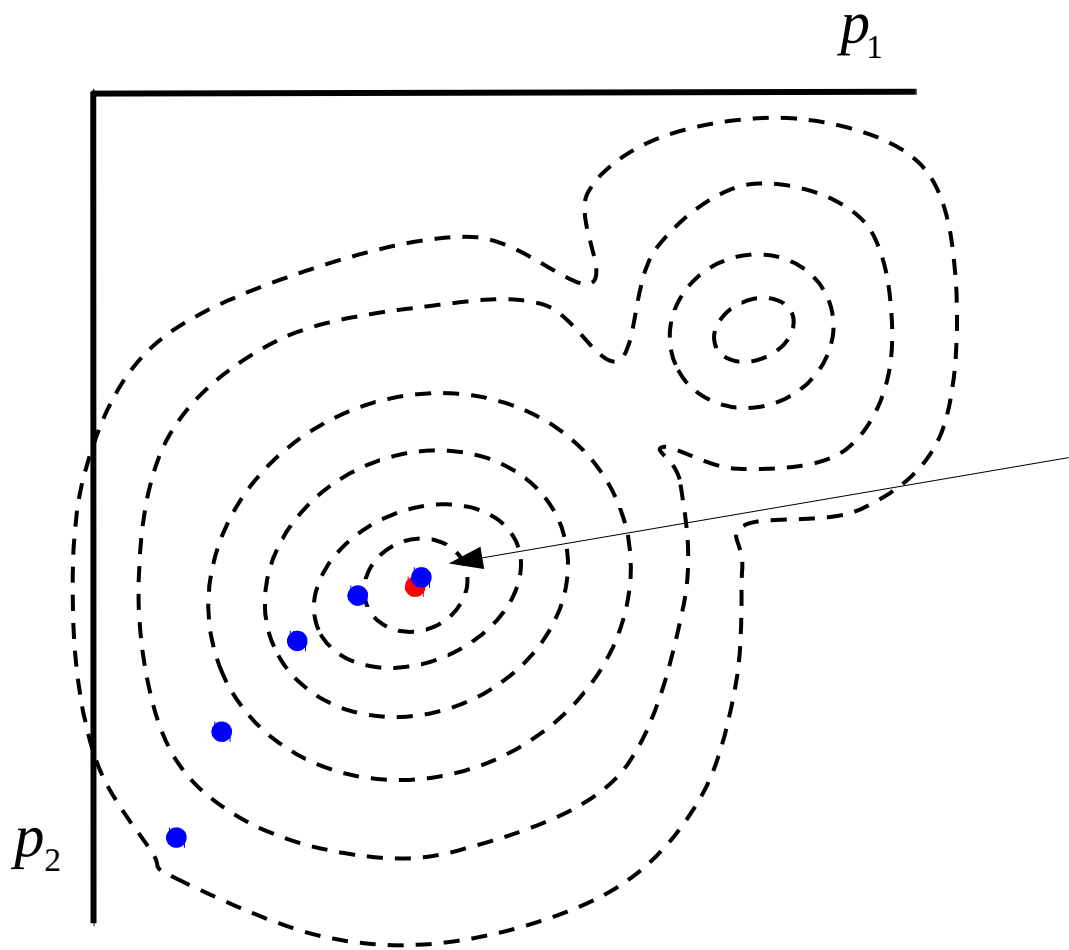




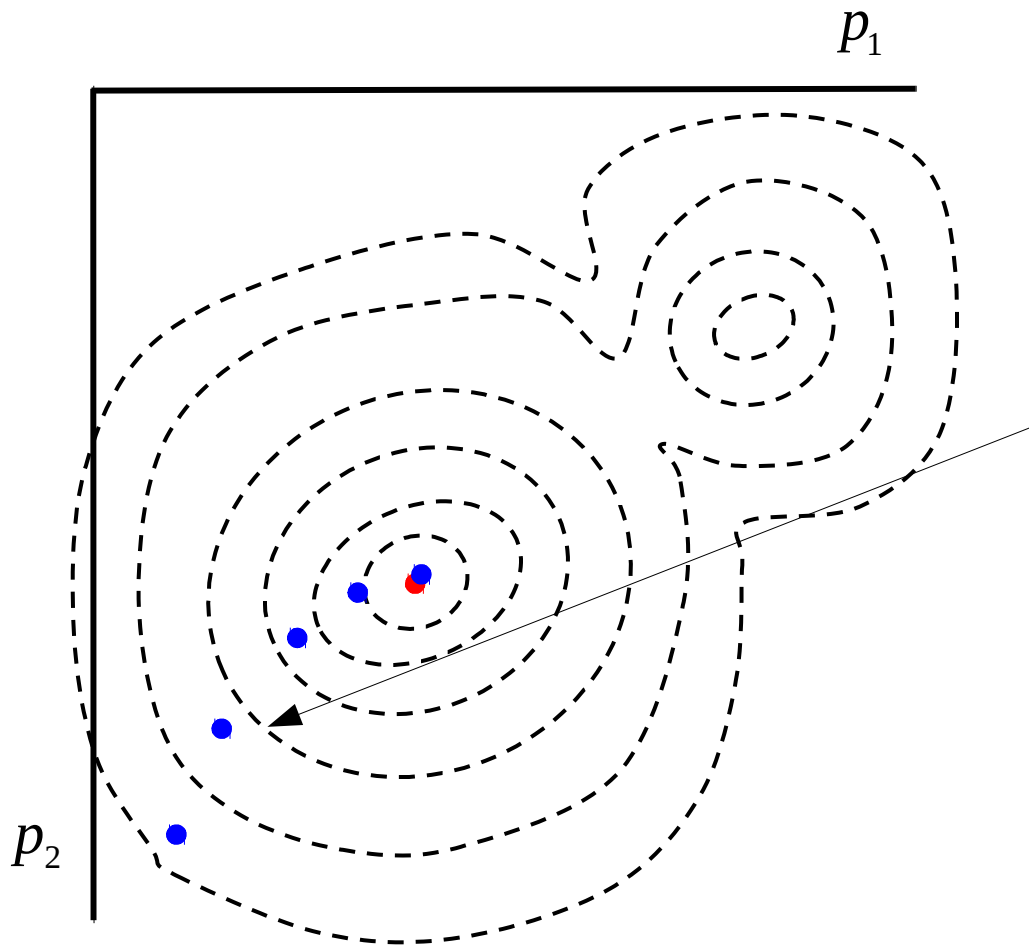




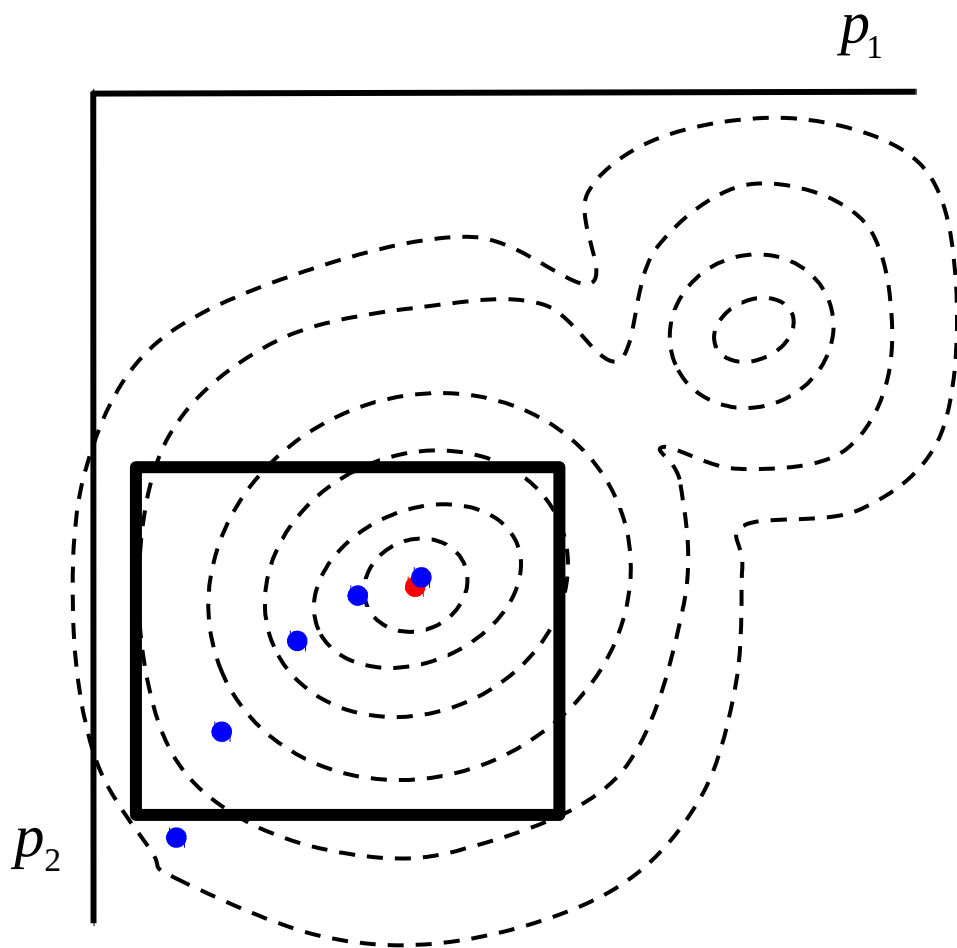


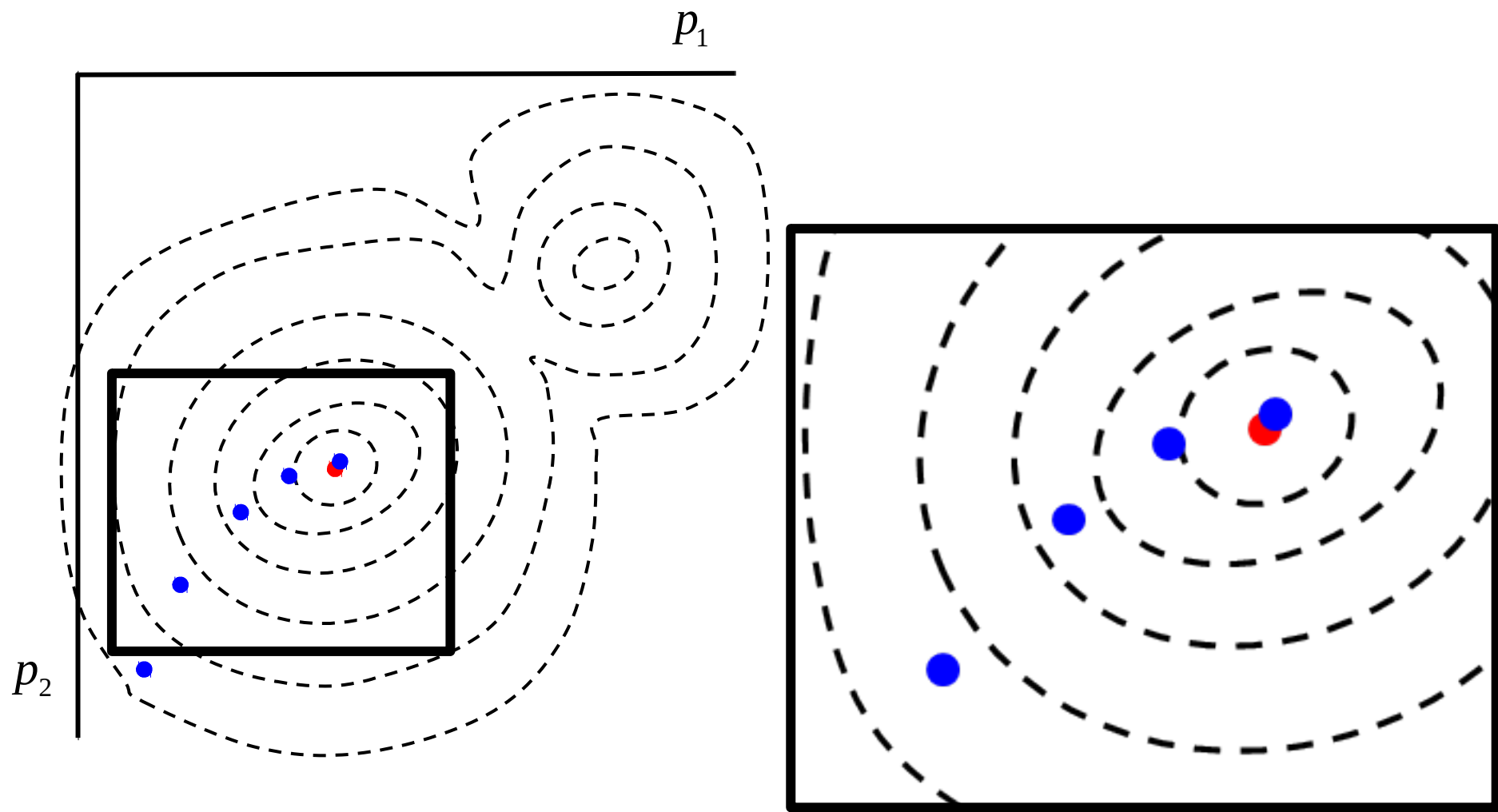


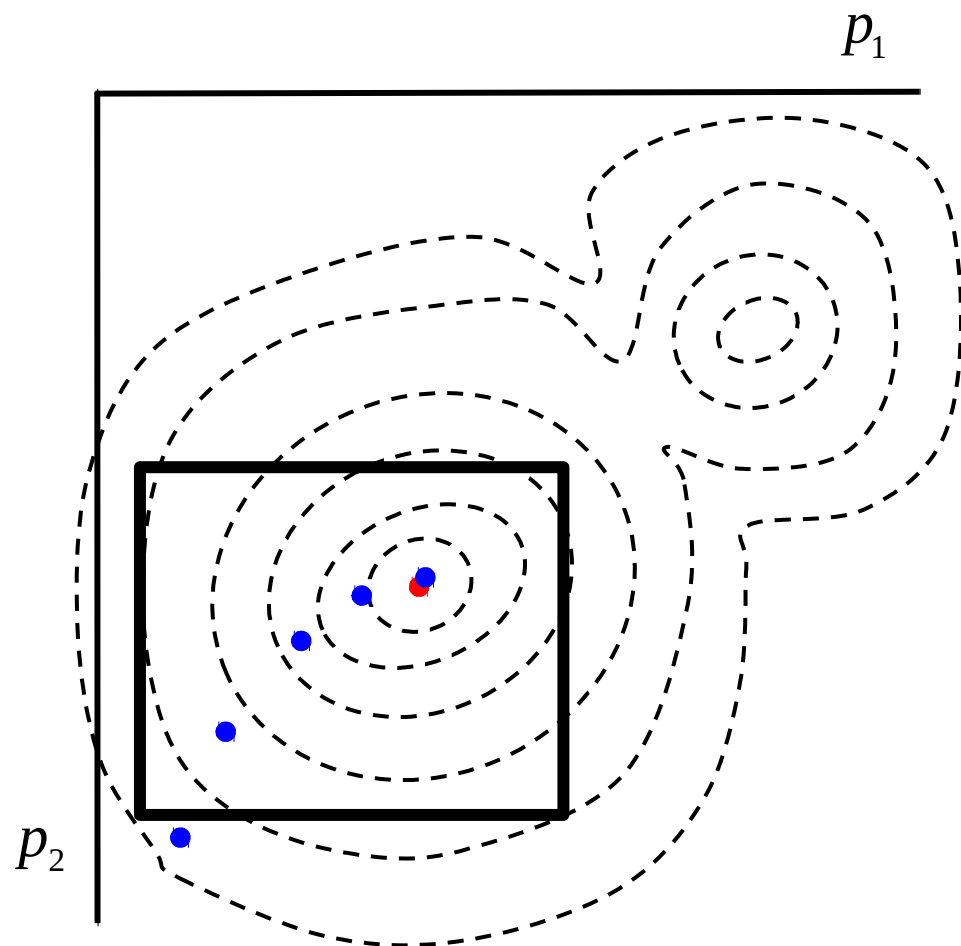
Estimated
minimum



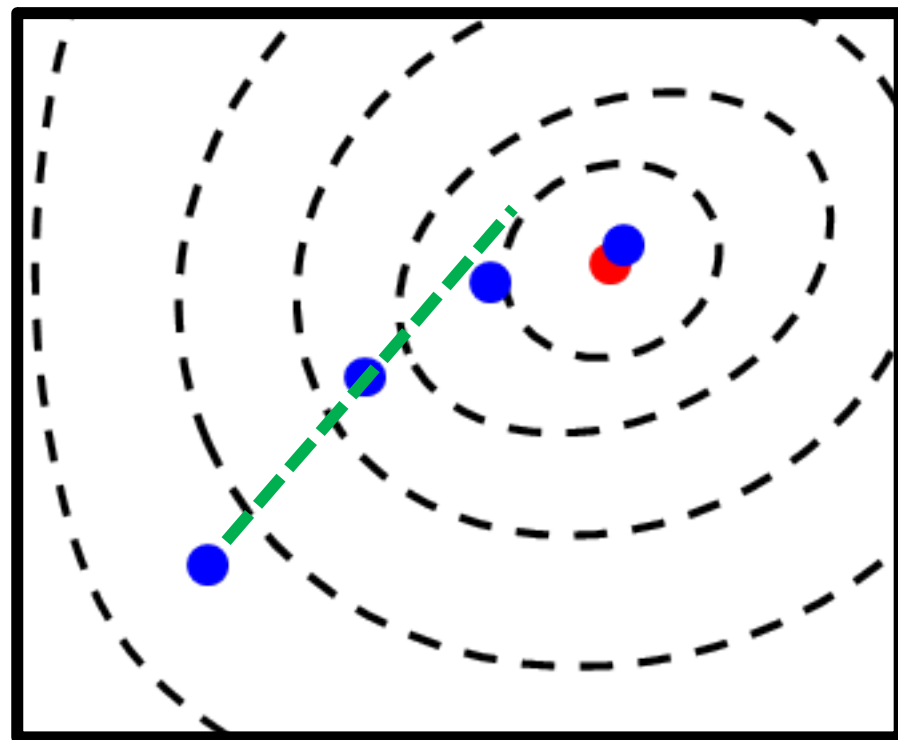
Given a point, it
is needed to
define a
direction and
the step length

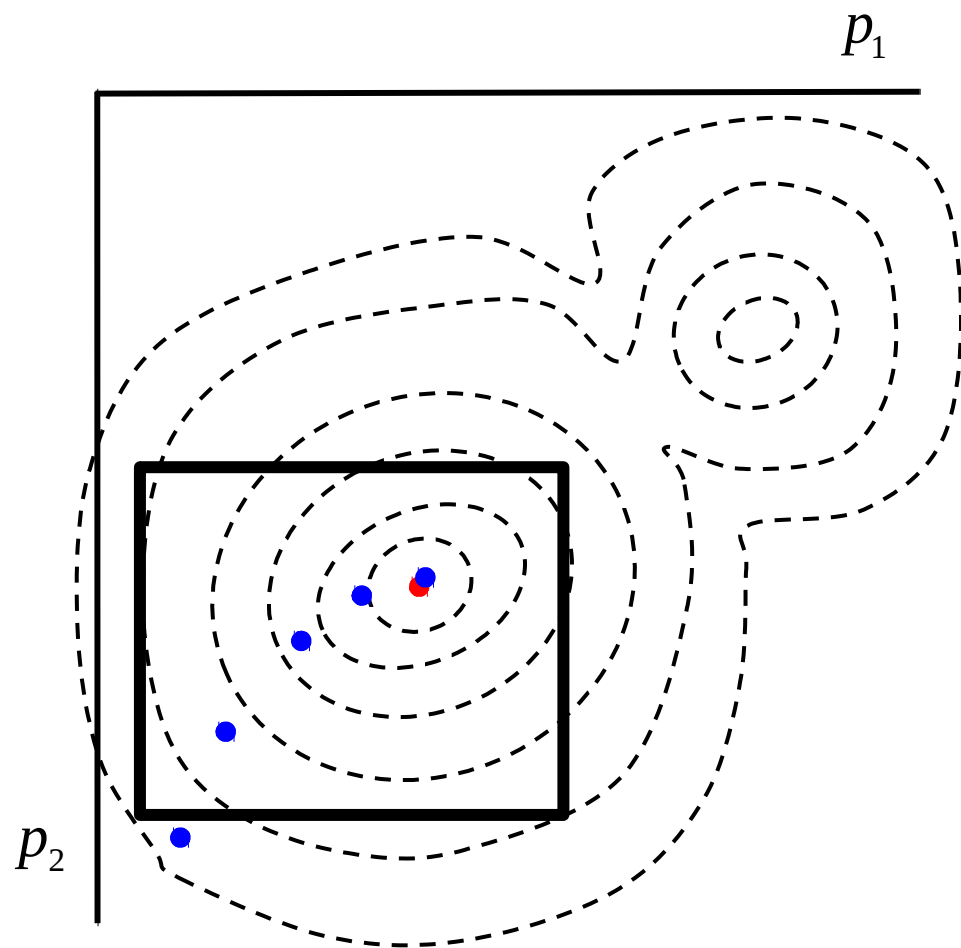




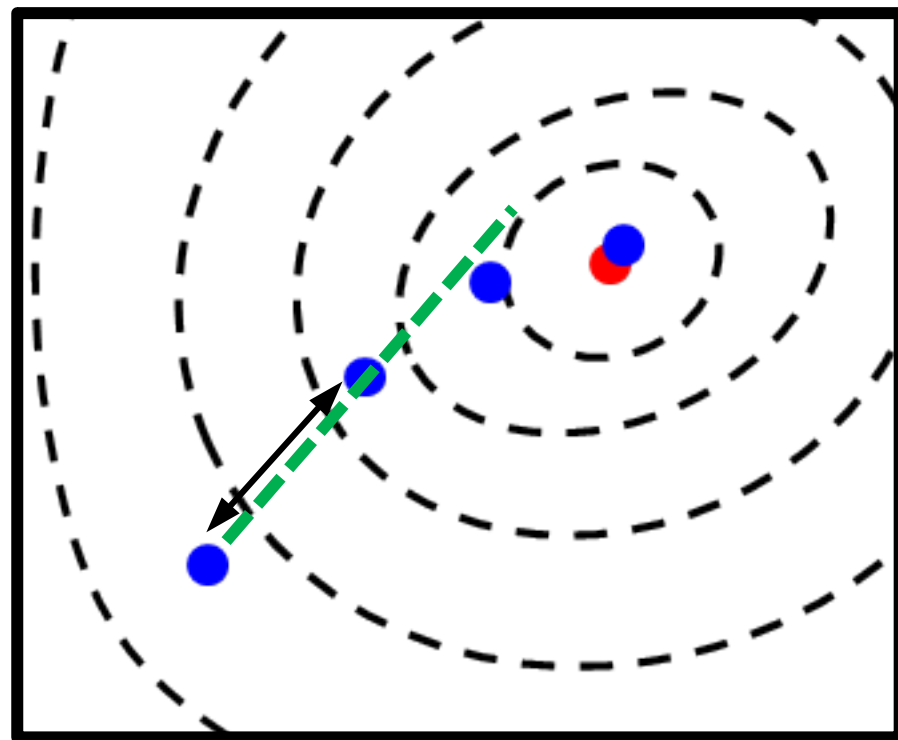


Direction

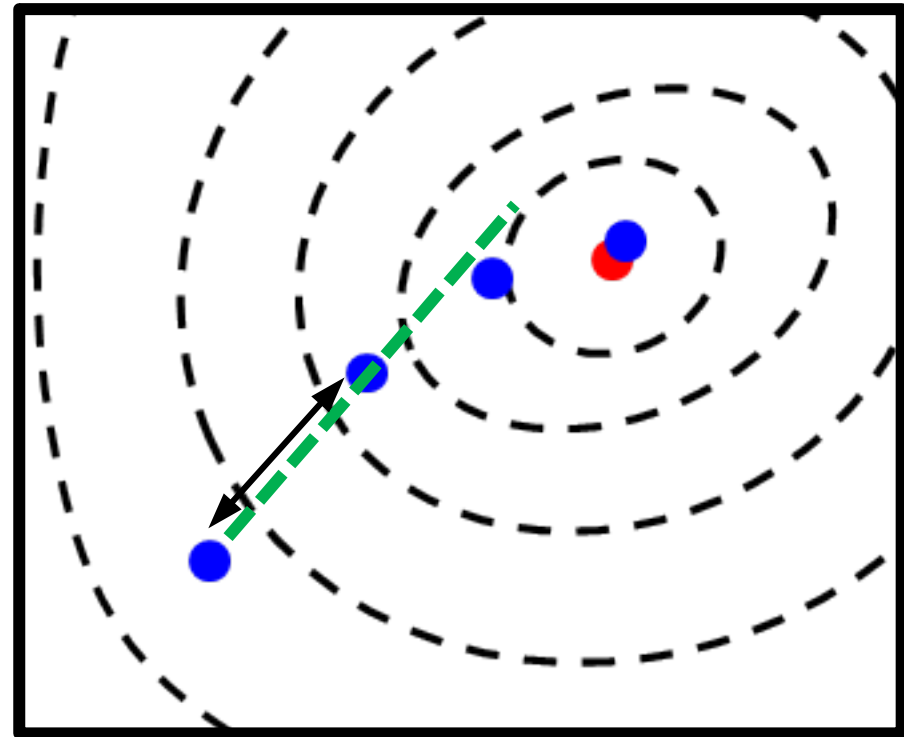




Step length



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



$$\Omega(\overline{p})$$

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\nabla}\bar{\nabla}\Omega(\bar{p}_0) \Delta\bar{p}$$

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\bar{\nabla}}\Omega(\bar{p}_0) \Delta\bar{p}$$

$$\bar{\bar{\nabla}}\Omega(\bar{p}_0) \Delta\bar{p} = -\bar{\nabla}\Omega(\bar{p}_0)$$

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \boxed{\bar{\nabla}^2\Omega(\bar{p}_0)} \Delta\bar{p}$$

$$\boxed{\bar{\nabla}^2\Omega(\bar{p}_0)} \Delta\bar{p} = -\bar{\nabla}\Omega(\bar{p}_0)$$

Difference between
the methods

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\nabla}^2\Omega(\bar{p}_0) \Delta\bar{p}$$

$$\bar{\nabla}^2\Omega(\bar{p}_0) \Delta\bar{p} = -\bar{\nabla}\Omega(\bar{p}_0)$$

Newton

$$\bar{\nabla}^2\Omega(\bar{p}_0)$$

Gauss - Newton

$$\bar{M}(\bar{p}_0)$$

Steepest decent

$$1/\eta$$

Levenberg -
Marquardt

$$\bar{M}(\bar{p}_0) + \lambda \bar{I}$$

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