Warm-Starting Strategies for Interior-Point Methods

Victor M. Zavala
Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

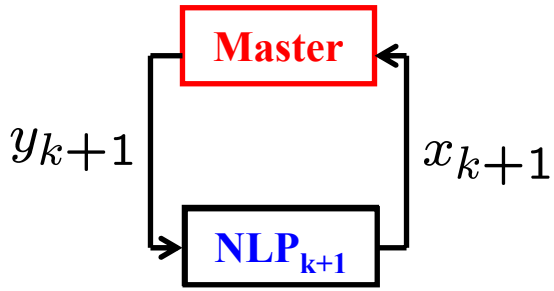
PSE Seminar
November 3rd 2006

Motivation

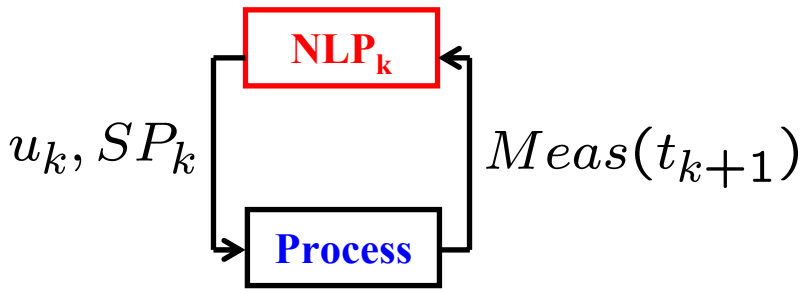
Motivation

Many Applications involve continuous solutions of *similar* NLPs

MINLP



Real-Time Optimization



Premise: Solution of NLP_k *might be* close to the Solution of NLP_{k+1}

How to re-use information efficiently to warm-start NLPs?

Active-Set Methods vs. Interior-Point Methods

Active-Set Methods

Efficient Warm-Starts - Active-Set Strategy of NLP Algorithm

NLP$_k$

$$\begin{aligned} \min \quad & f(x, p_k) \\ \text{s.t.} \quad & c(x, p_k) = 0 \\ & x \geq 0 \end{aligned}$$

KKT Conditions of NLP$_k$

$$\begin{aligned} \nabla f(x, p_k) + \nabla c(x, p_k)\,\lambda - \nu &= 0 \\ c(x, p_k) &= 0 \\ XVe &= 0 \\ x, \nu &\geq 0 \end{aligned}$$

Disadvantages

Exponential Complexity -Poor Guess of A_k - Structure NOT Preserved Along Iterations

Active-Set Methods

Guess $A = \{\, i \,|\, x^{(i)} = 0 \}$

Solve Subproblem

$$\begin{aligned} \min \quad & f(x, p_k) \\ \text{s.t.} \quad & c(x, p_k) = 0 \\ & x^{(i)} = 0, \; i \in A \end{aligned}$$

Check KKT of NLP$_k$

Re-guess $A = \{\, i \,|\, x^{(i)} = 0 \}$

$$s_k^* = [x_k^*, \lambda_k^*, \nu_k^*]$$

$$A_k = \{\, i \,|\, x_k^{*(i)} = 0 \}$$

Active-Set Methods

Advantage - Optimal Solution of NLP_k *USUALLY* a Good Starting Point for NLP_{k+1}

Nominal Problem - NLP_k

$$\begin{aligned} \min \quad & f(x, p_k) \\ \text{s.t.} \quad & c(x, p_k) = 0 \\ & x \geq 0 \end{aligned}$$

$$\downarrow$$

$$s_k^*$$

$$A_k = \{\, i \,|\, x_k^{*(i)} = 0 \}$$

$$\longrightarrow$$

Perturbed Problem - NLP_{k+1}

$$\begin{aligned} \min \quad & f(x, p_{k+1}) \\ \text{s.t.} \quad & c(x, p_{k+1}) = 0 \\ & x \geq 0 \end{aligned}$$

$$\downarrow$$

$$s_{k+1}^*$$

$$A_{k+1} = \{\, i \,|\, x_{k+1}^{*(i)} = 0 \}$$

IF A_k **and** A_{k+1} **differ in a few variables**

THEN s_k^* **excellent warm-starting point for NLP_{k+1} -Few Iterations-**

ELSE Combinatorial Problem - Difficult to Predict

Interior-Point Methods

NLP$_k$

$$\begin{aligned} \min \quad & f(x, p_k) \\ \text{s.t.} \quad & c(x, p_k) = 0 \\ & x \geq 0 \end{aligned}$$

$\longrightarrow$

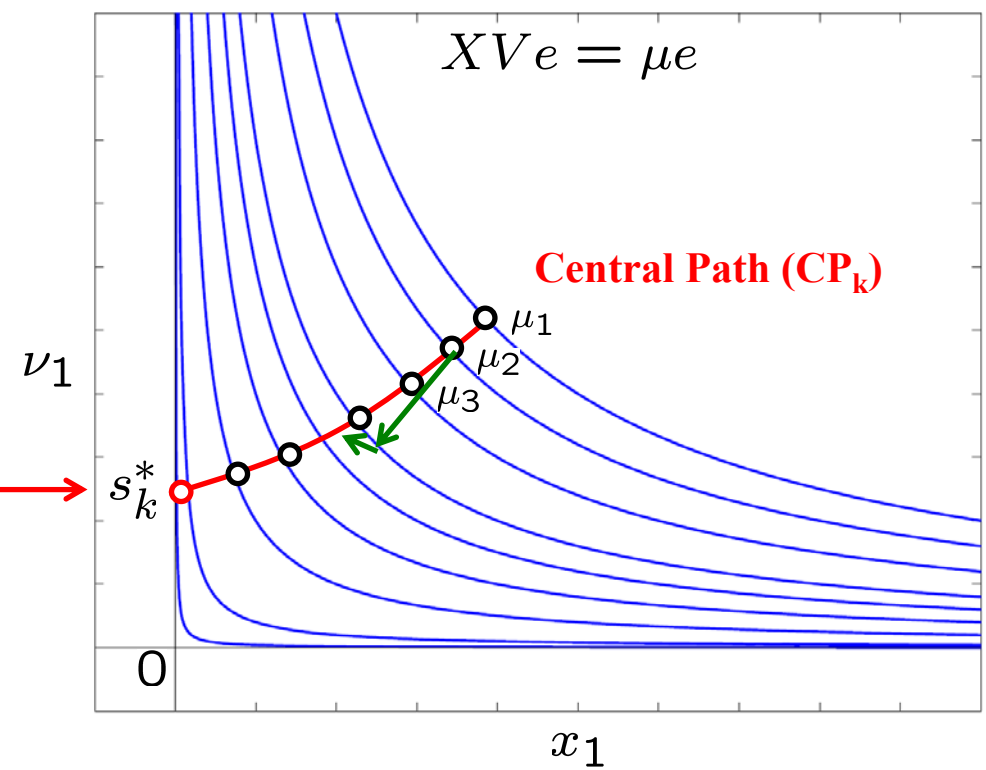
BP$_{k,\ell}$

$$\begin{aligned} \min \quad & f(x, p_k) - \mu_\ell \sum_{i=1}^{nx} \ln\left(x^{(i)}\right) \\ \text{s.t.} \quad & c(x, p_k) = 0 \end{aligned}$$

KKT Conditions of BP$_{k,\ell}$

$$\begin{aligned} \nabla f(x, p_k) + \nabla c(x, p_k)\,\lambda - \nu &= 0 \\ c(x, p_k) &= 0 \\ XVe &= \mu_\ell e \\ x, z &\geq 0 \end{aligned}$$

$\mu = 1 \times 10^{-6}$



Polynomial Complexity

Interior-Point Methods

Infeasible-Path Methods Solve Sequence of BPs driving $\mu_\ell \to 0$

KKT Conditions of $BP_{k,\ell}$

$$\begin{aligned} \nabla f(x,p_k) + \nabla c(x,p_k)\,\lambda - \nu &= 0 \\ c(x,p_k) &= 0 \\ XVe &= \mu_\ell e \\ \boxed{x,\nu \geq 0} \end{aligned}$$

Newton's Method - KKT System (Fixed Structure)

$$\begin{bmatrix} W(x_j,\lambda_j) & A(x_j) & -I \\ A(x_j)^T & 0 & 0 \\ V_j & 0 & X_j \end{bmatrix} \boxed{\begin{bmatrix} \Delta x_j \\ \Delta\lambda_j \\ \Delta\nu_j \end{bmatrix}} = -\begin{bmatrix} \nabla f(x_j) + A(x_j)\lambda_k - \nu_j \\ c(x_j) \\ X_j V_j e - \mu_\ell e \end{bmatrix}$$

Non-negativity of x and z is imposed *explicitly* at every iteration

Fraction-to-the-boundary

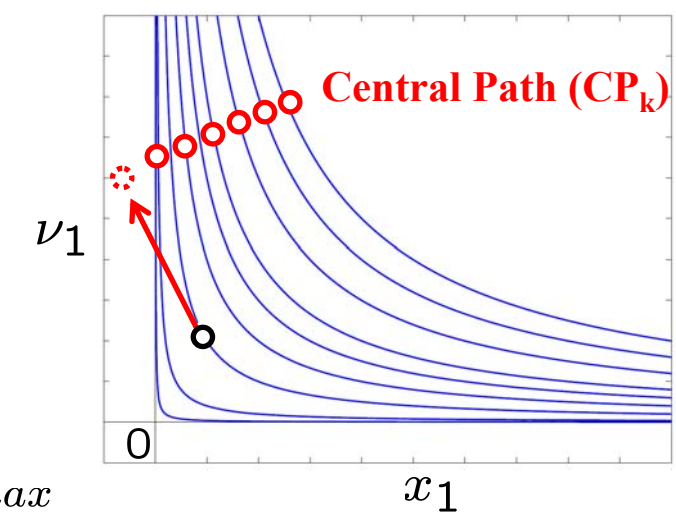
Maximum Stepsize - Stay Inside of Feasible Region

$$\alpha_j^{max} = \max\{\alpha \in (0,1] : x_j + \alpha d_j^x \geq (1-\tau)x_j\}$$

$$\alpha_j^{\nu} = \max\{\alpha \in (0,1] : \nu_j + \alpha d_j^{\nu} \geq (1-\tau)\nu_j\}$$

$$x_{j+1} = x_j + \alpha_j^{max}\Delta x_j$$

$$\nu_{j+1} = \nu_j + \alpha_j^{\nu}\Delta\nu_j$$



Final stepsize obtained by backtracking line-search from α_j^{max}

Interior-Point Methods

Warm-Starts - Optimal Solution of NLP$_k$ *SOMETIMES* a Good Starting Point

NLP$_k$

$$\begin{aligned} \min \quad & f(x, p_k) \\ \text{s.t.} \quad & c(x, p_k) = 0 \\ & x \geq 0 \end{aligned}$$

$\longrightarrow$

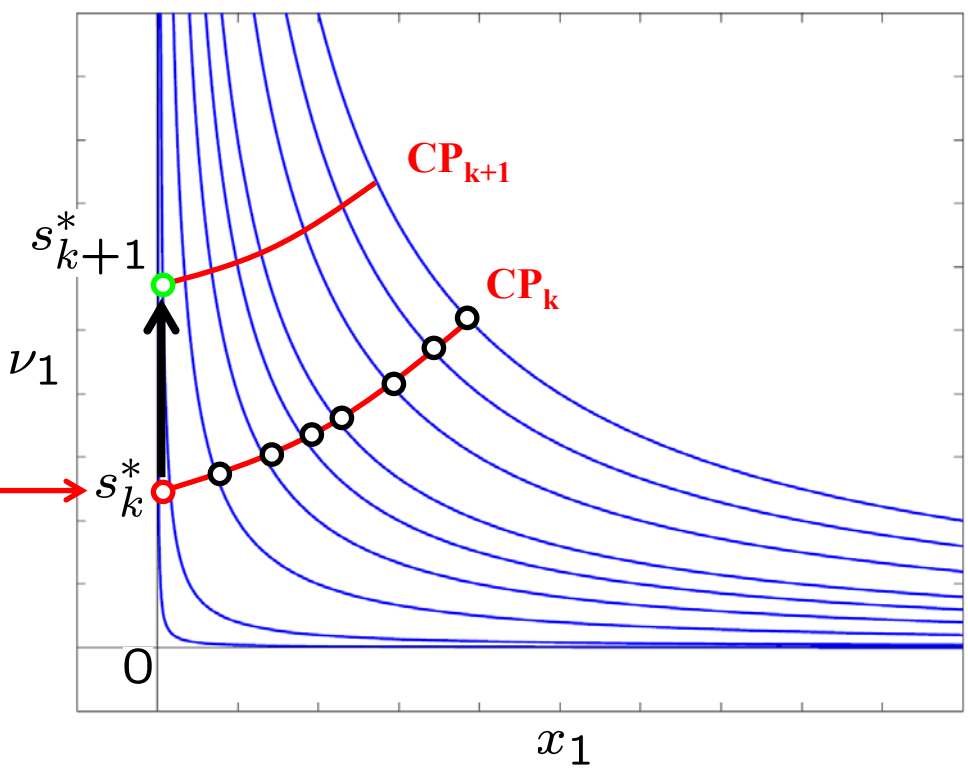
NLP$_{k+1}$

$$\begin{aligned} \min \quad & f(x, p_{k+1}) \\ \text{s.t.} \quad & c(x, p_{k+1}) = 0 \\ & x \geq 0 \end{aligned}$$

Case 1: $\|p_{k+1} - p_k\|$ **Does NOT Impose Change in Active Set** $A_{k+1} = A_k$

Is it possible to precompute s^*_{k+1} **?**

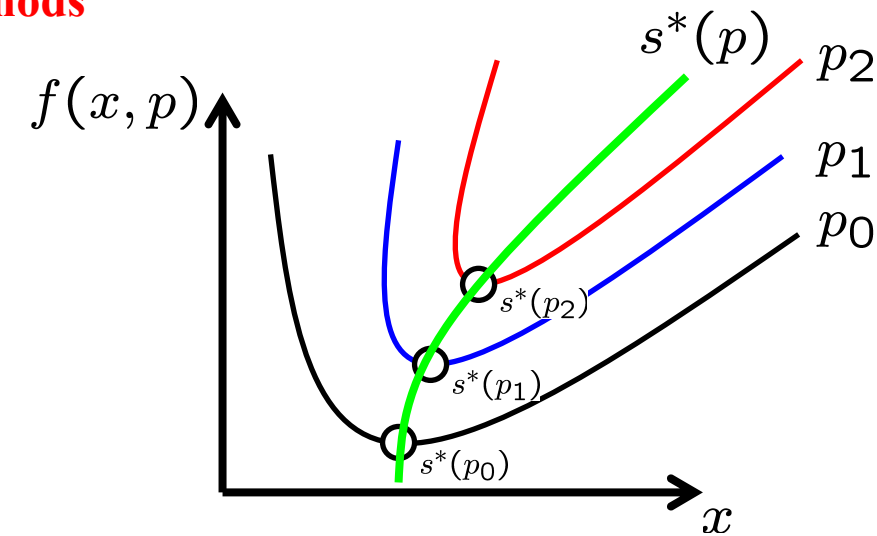
$\mu = 1 \times 10^{-6}$



NLP Sensitivity

Apply to Active-Set and Interior-Point Methods

$$\left.\begin{array}{ll} \min & f(x,p) \\ \text{s.t.} & c(x,p)=0 \\ & x \geq 0 \end{array}\right\} \mathrm{P}(p)$$



Solution Triplet

$$s^*(p)^T = [x^{*T}\ \lambda^{*T}\ \nu^{*T}]$$

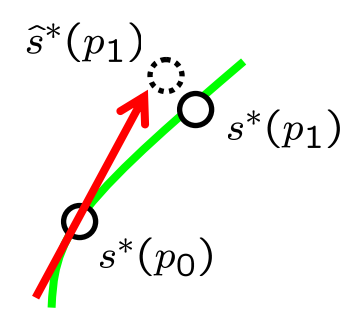
Optimality Conditions of $\mathrm{P}(p)$

$$\begin{aligned} \nabla_x f(x,p) + \nabla_x c(x,p)\,\lambda - \nu &= 0 \\ c(x,p) &= 0 \\ XVe &= 0 \end{aligned}$$

NLP Sensitivity → Existence and Differentiability of Path $s^*(p)$ **→ LICQ, SSOC, SC**

→ Main Idea: Obtain $\left.\dfrac{\partial s}{\partial p}\right|_{p_0}$ **and find** $\hat{s}^*(p_1)$ **by Taylor Series Expansion**

$$\hat{s}^*(p_1) \approx s^*(p_0) + \left.\frac{\partial s}{\partial p}^T\right|_{p_0} (p - p_0)$$



NLP Sensitivity

Obtaining $\left.\frac{\partial s}{\partial p}\right|_{p_0}$

Optimality Conditions of $\mathrm{P}(p)$

$$\left.\begin{aligned} \nabla_x \mathcal{L} = \nabla_x f(x,p) + \nabla_x c(x,p)\,\lambda - \nu &= 0 \\ c(x,p) &= 0 \\ XVe &= 0 \end{aligned}\right\} \quad \mathbf{F}(s,p) = 0$$

Apply Implicit Function Theorem to $\mathbf{F}(s,p) = 0$ **around** $(p_0, s^*(p_0))$

$$\frac{\partial \mathbf{F}(s^*(p_0),p_0)}{\partial s} \left.\frac{\partial s}{\partial p}\right|_{p_0} + \frac{\partial \mathbf{F}(s^*(p_0),p_0)}{\partial p} = 0$$

$$\begin{bmatrix} W(s^*(p_0)) & A(x^*(p_0)) & -I \\ A(x^*(p_0))^T & 0 & 0 \\ V^*(p_0) & 0 & X^*(p_0) \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial p} \\ \frac{\partial \lambda}{\partial p} \\ \frac{\partial \nu}{\partial p} \end{bmatrix} + \begin{bmatrix} \nabla_{x,p}\mathcal{L}(s^*(p_0)) \\ \nabla_p\, c(x^*(p_0)) \\ 0 \end{bmatrix} = 0$$

KKT Matrix

$$\begin{bmatrix} W(x_k,\lambda_k) & A(x_k) & -I \\ A(x_k)^T & 0 & 0 \\ V_k & 0 & X_k \end{bmatrix}$$

- **Already Factored at Solution**
- **FAST Sensitivity Calculation**
- **Approximation Preserves Active Set of Nominal Problem**

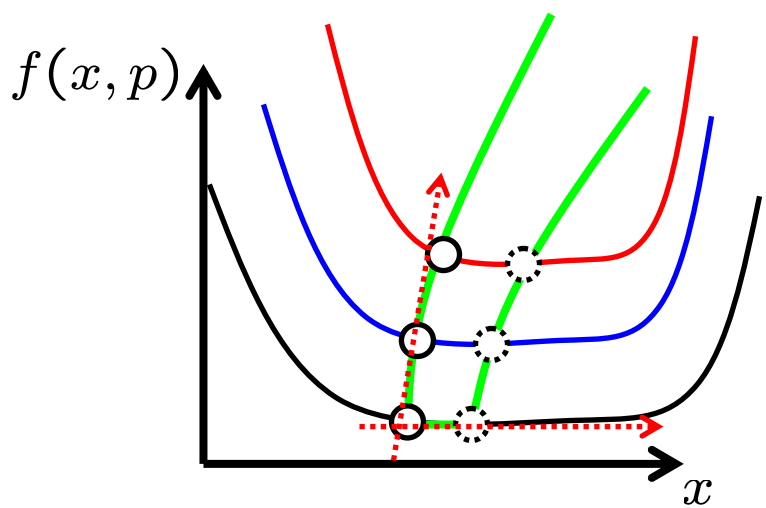
NLP Sensitivity

Degenerate Nominal Solutions

Nonstrict Complementarity, LICQ - KKT Matrix Singular

$$\begin{bmatrix} W(s^*(p_0)) & A(x^*(p_0)) & -I \\ A(x^*(p_0))^T & 0 & 0 \\ \cancel{V^*(p_0)} & \cancel{0} & \cancel{X^*(p_0)} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial p} \\ \frac{\partial \lambda}{\partial p} \\ \frac{\partial \nu}{\partial p} \end{bmatrix} + \begin{bmatrix} \nabla_{x,p}\mathcal{L}(s^*(p_0)) \\ \nabla_p\, c(x^*(p_0)) \\ 0 \end{bmatrix} = 0$$

Strong Second-Order Conditions DO NOT Hold

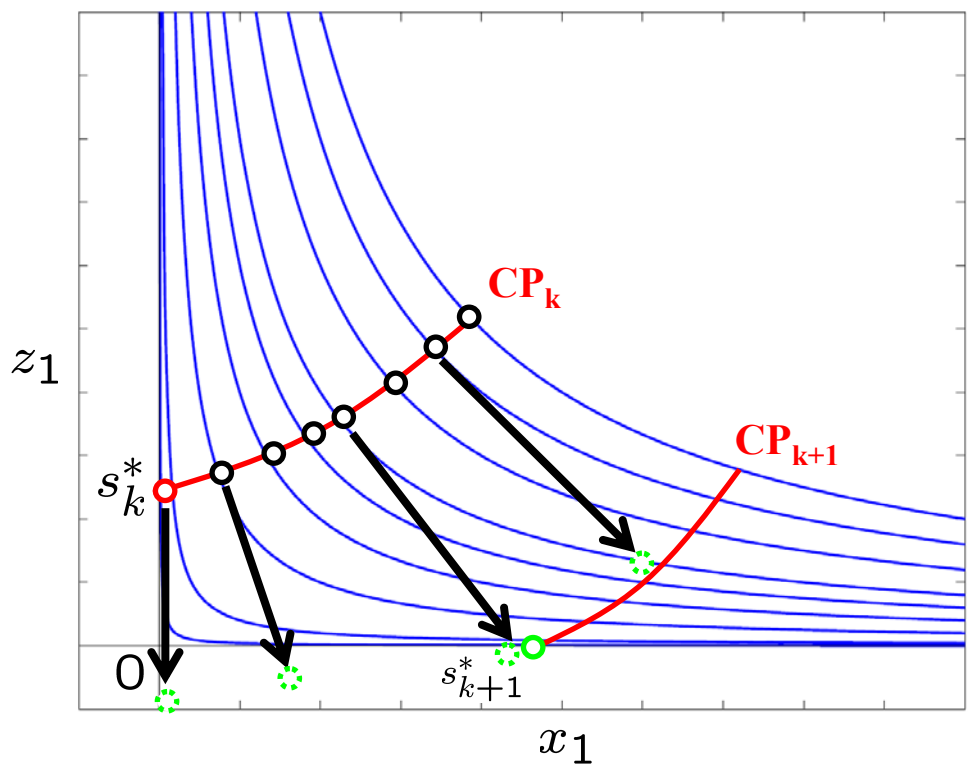


$$\left.\frac{\partial s}{\partial p}\right|_{p_0}$$ **Non-Unique**

Bifurcation Point - Continuity Exists but NOT Differentiability

Interior-Point Methods

Case 2: $\|p_{k+1} - p_k\|$ **Induce Change in Active Set** $A_{k+1} \neq A_k$



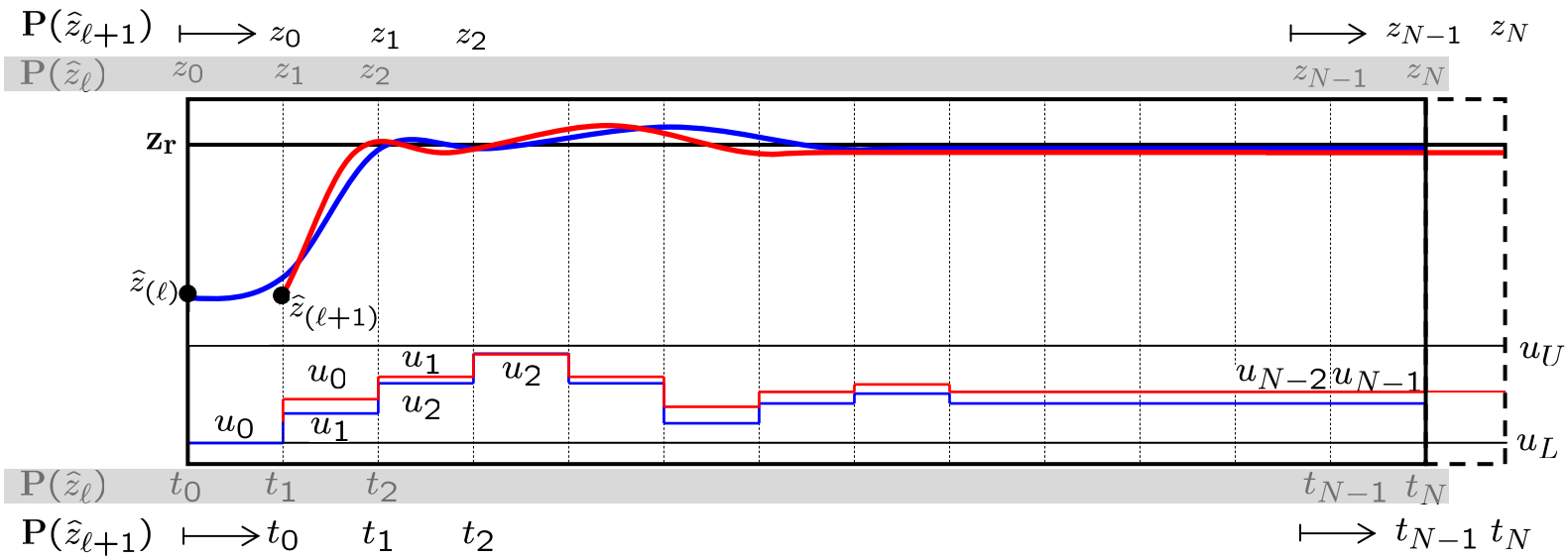
Initializing NLP$_{k+1}$ with s_k^* ***might not*** **decrease number of iterations**

Use solution of a barrier sub-problem $s_k^*(\mu)$ **(Yildrim & Wright, Gondzio & Grothey)**

Which One? Trial and Error, Very Large μ

Case Study – Nonlinear Model Predictive Control

Nonlinear Model Predictive Control



$$\mathbf{P}(\hat{z}_\ell)$$

$$\left.\begin{aligned}
\min \quad & \sum_{m=0}^{M-1}\left\{\sum_{k=0}^{K_m-1}\varphi_m^k\left(z_m^k, y_m^k, u_m^k, \pi_m, \Pi\right) + E_m\left(z_m^{K_m}, y_m^{K_m}, u_m^{K_m}, \pi_m, \Pi\right)\right\} \\
\text{s.t.} \quad & \\
z_m^{k+1} &= z_m^k + \mathbf{A}_m y_m^k \\
0 &= h_{m,k}\left(z_m^k, y_m^k, u_m^k, \pi_m, \Pi\right) \\
z_{m+1}^0 &= \bar{f}_m\left(z_m^{K_m}, y_m^{K_m}, u_m^{K_m}\right) \\
\boxed{z_0^0 = \hat{z}(\ell) = p_0} & \\
m &= 0 \ldots\ldots M-1 \qquad k = 0 \ldots\ldots K_m - 1
\end{aligned}\right\} \mathbf{P}(p_0)$$

NLPs are Parametric

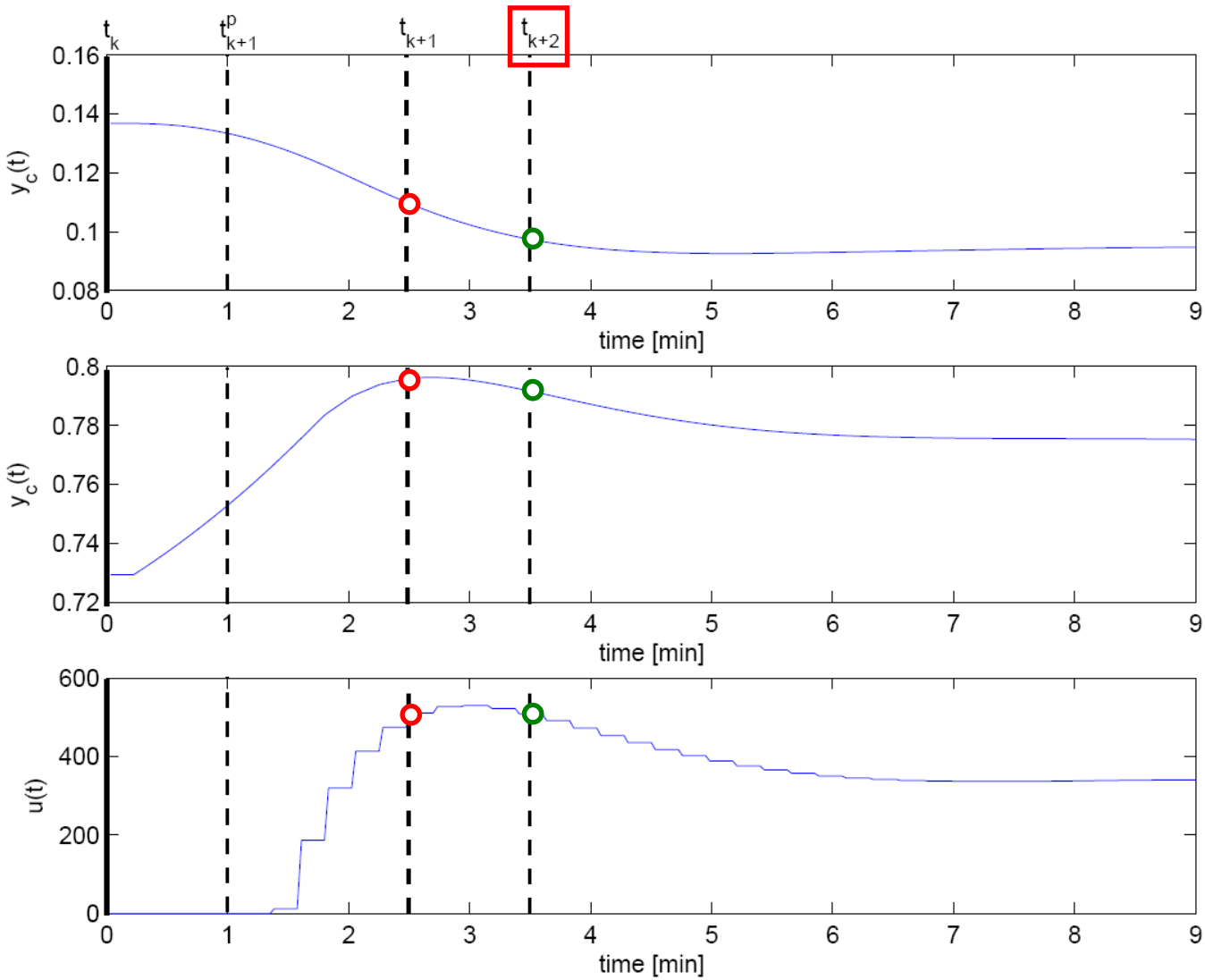
$$\mathbf{P}(p)$$

$$\mathbf{P}(\hat{z}_{\ell+1})$$

$$\begin{aligned}
\min \quad & \sum_{m=0}^{M-1}\left\{\sum_{k=0}^{K_m-1}\varphi_m^k\left(z_m^k, y_m^k, u_m^k, \pi_m, \Pi\right) + E_m\left(z_m^{K_m}, y_m^{K_m}, u_m^{K_m}, \pi_m, \Pi\right)\right\} \\
\text{s.t.} \quad & \\
z_m^{k+1} &= z_m^k + \mathbf{A}_m y_m^k \\
0 &= h_{m,k}\left(z_m^k, y_m^k, u_m^k, \pi_m, \Pi\right) \\
z_{m+1}^0 &= \bar{f}_m\left(z_m^{K_m}, y_m^{K_m}, u_m^{K_m}\right) \\
\boxed{z_0^0 = \hat{z}(\ell+1) = p} & \\
m &= 0 \ldots\ldots M-1 \qquad k = 0 \ldots\ldots K_m - 1
\end{aligned}$$

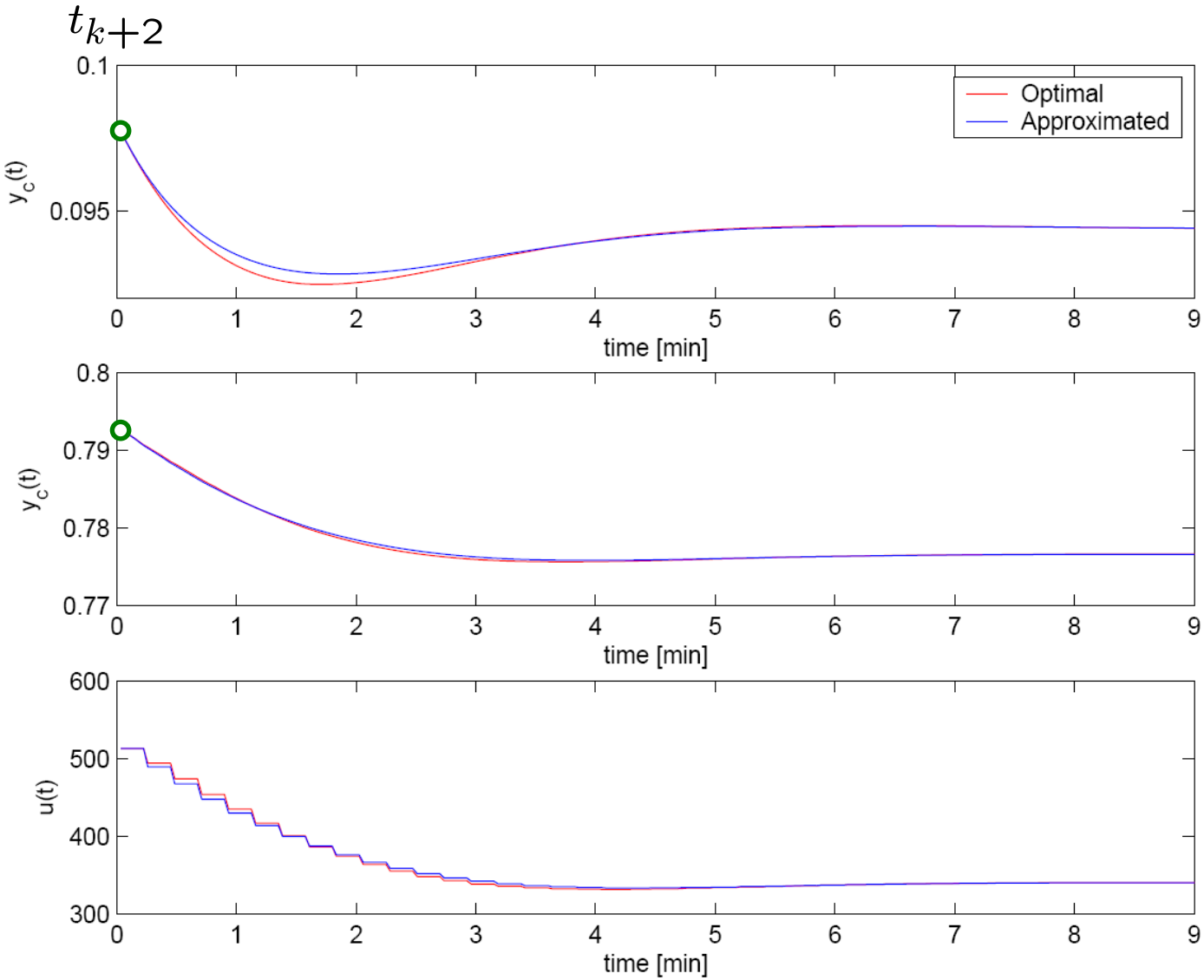
Nonlinear Model Predictive Control

How to Warm-Start Neighboring NLPs? - NLP Sensitivity



Nonlinear Model Predictive Control

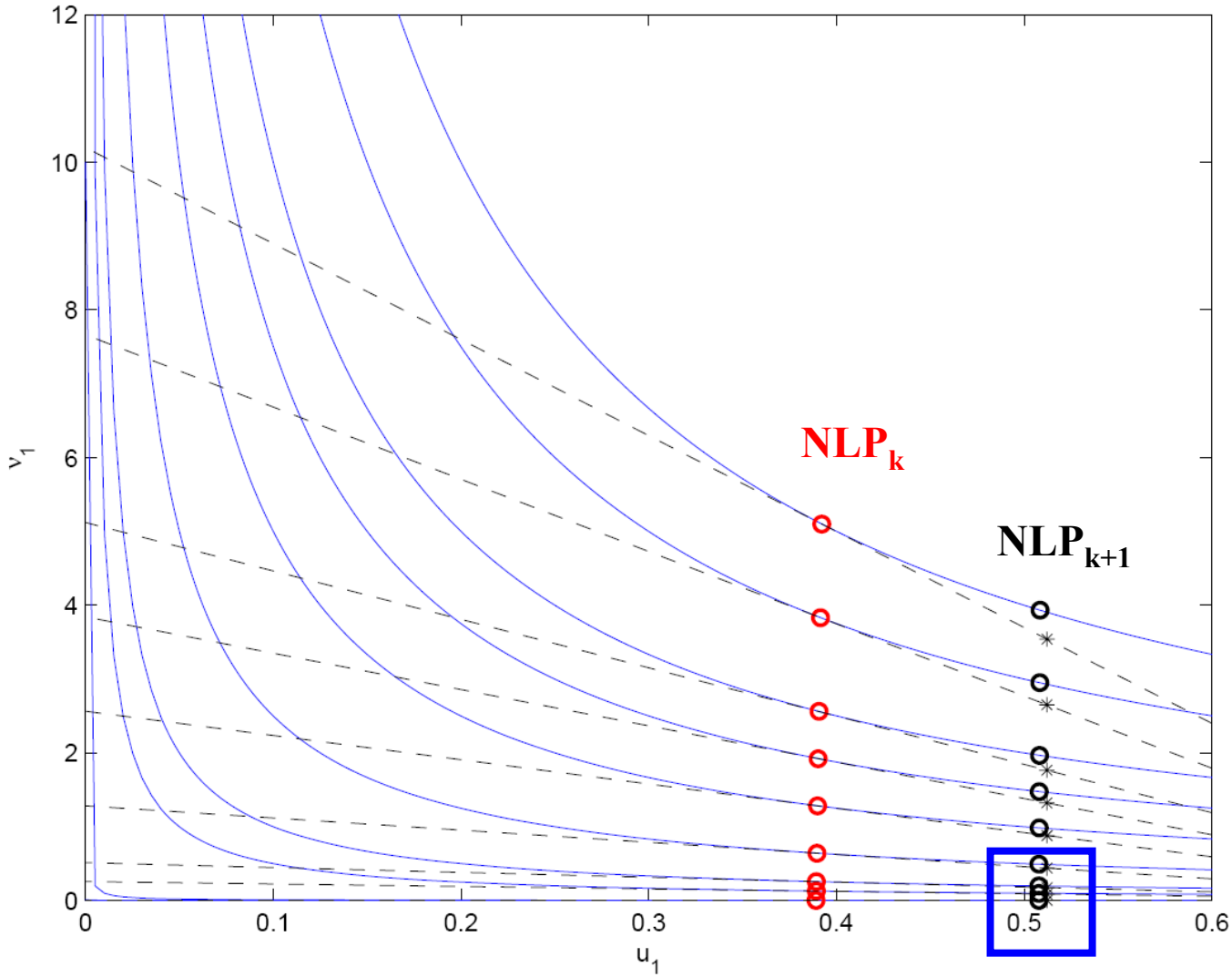
Case 1: No Change in the Active-Set



Nonlinear Model Predictive Control

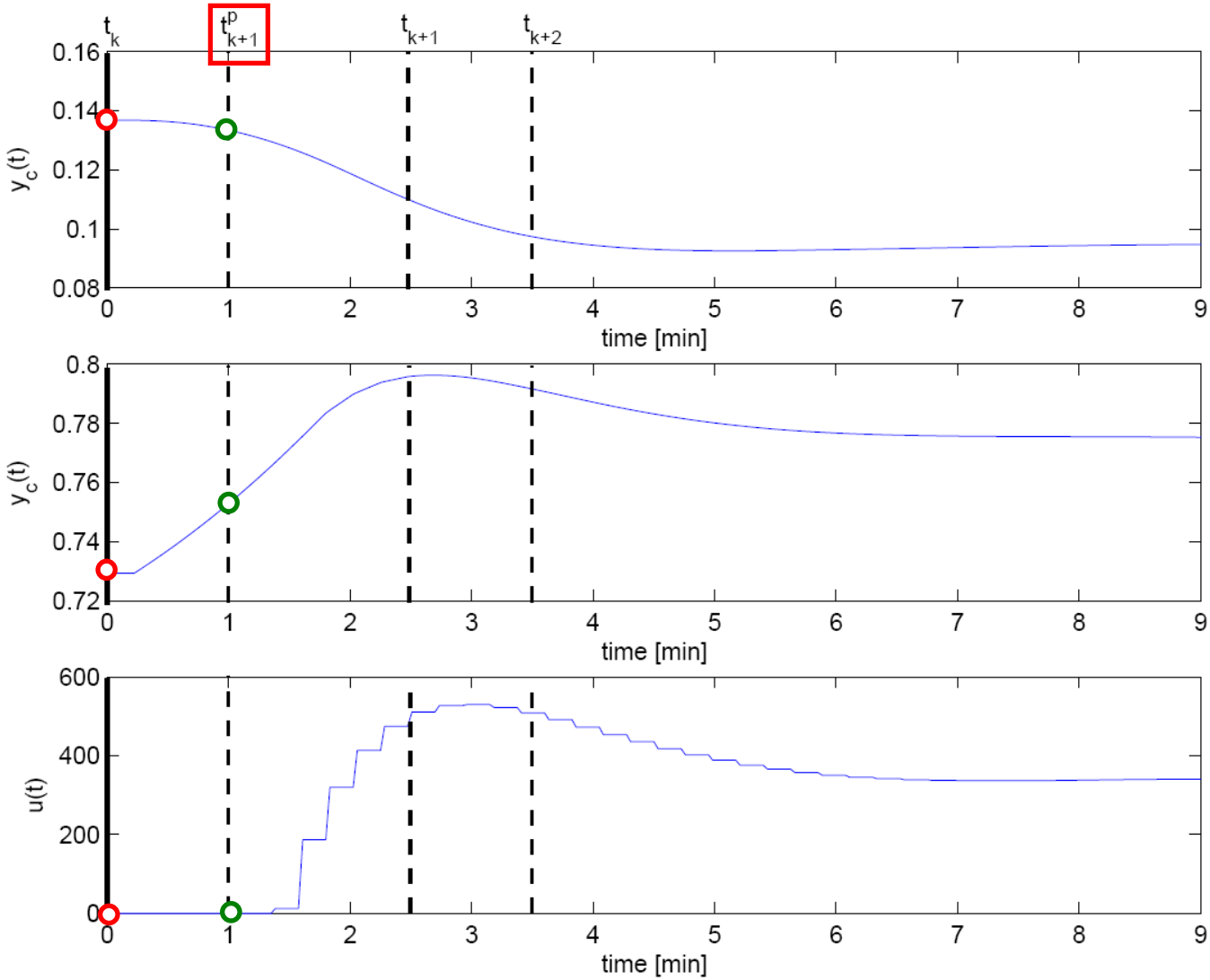
Case 1: No Change in the Active-Set

NLP Sensitivity Provides Excellent Warm-Starting Point

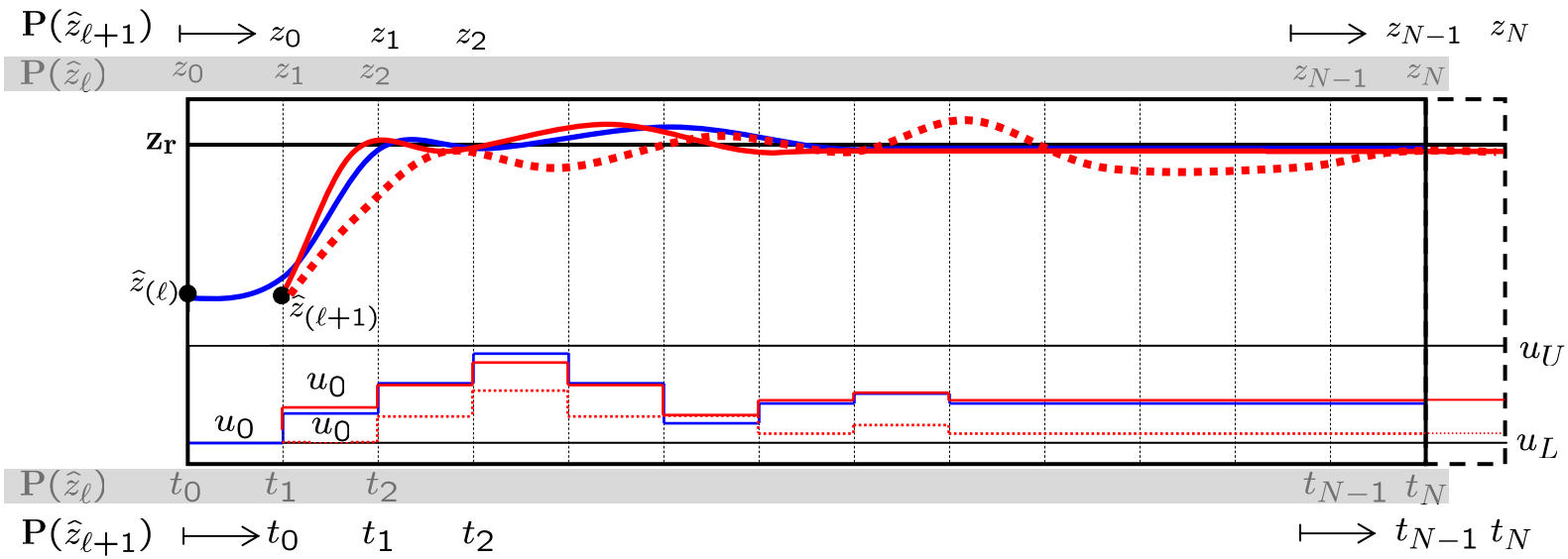


Nonlinear Model Predictive Control

Case 2: Change in the Active-Set



Nonlinear Model Predictive Control



Optimal Solution $P(\hat{z}_\ell)$ → $u_0 = u_L$

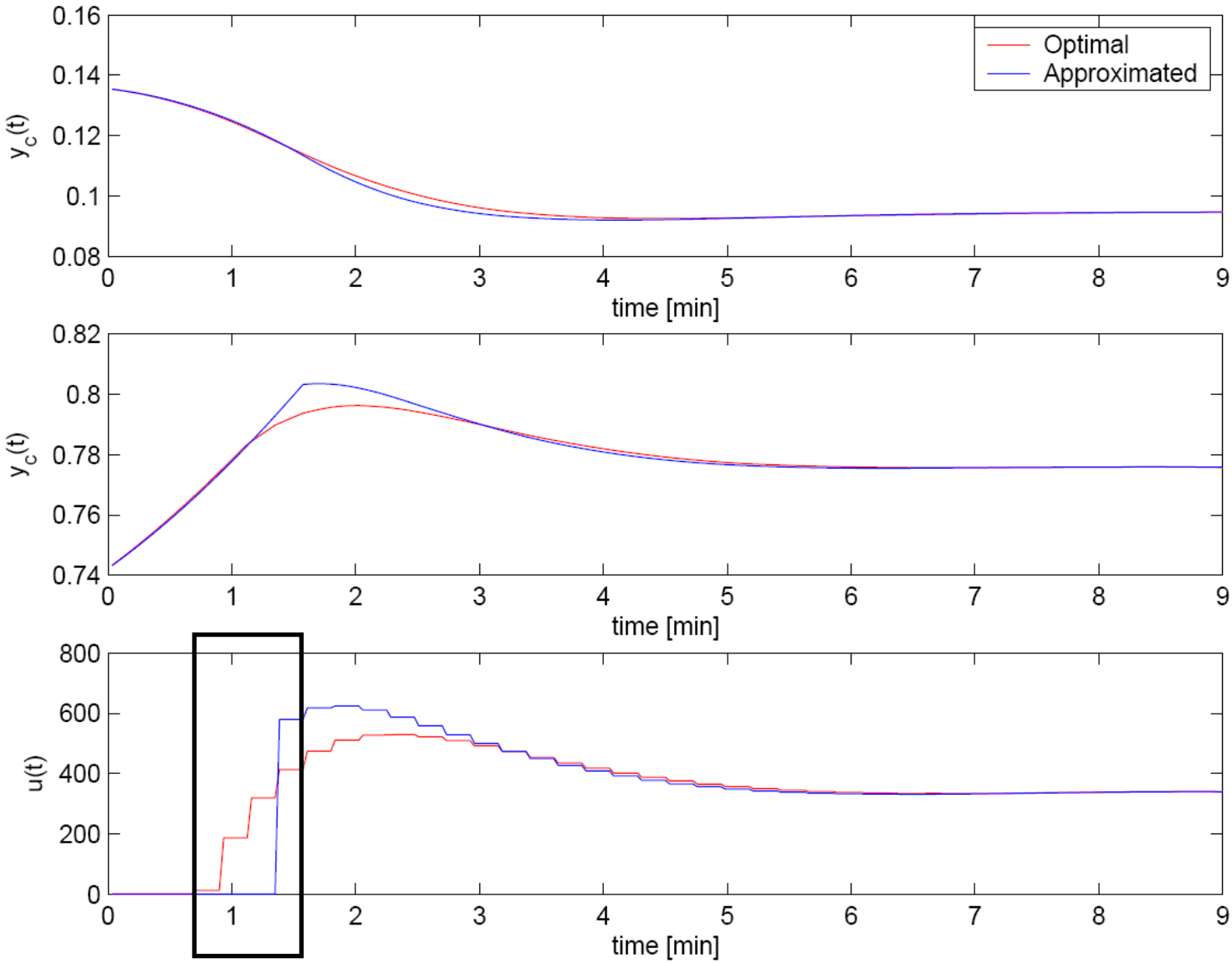
Optimal Solution $P(\hat{z}_{\ell+1})$ → $u_0 > u_L$

Approximate $P(\hat{z}_{\ell+1})$ **by Perturbing** $P(\hat{z}_\ell)$ ⇢ $u_0 = u_L$

Inconsistent Approximation → Inconsistent Active-Sets

Nonlinear Model Predictive Control

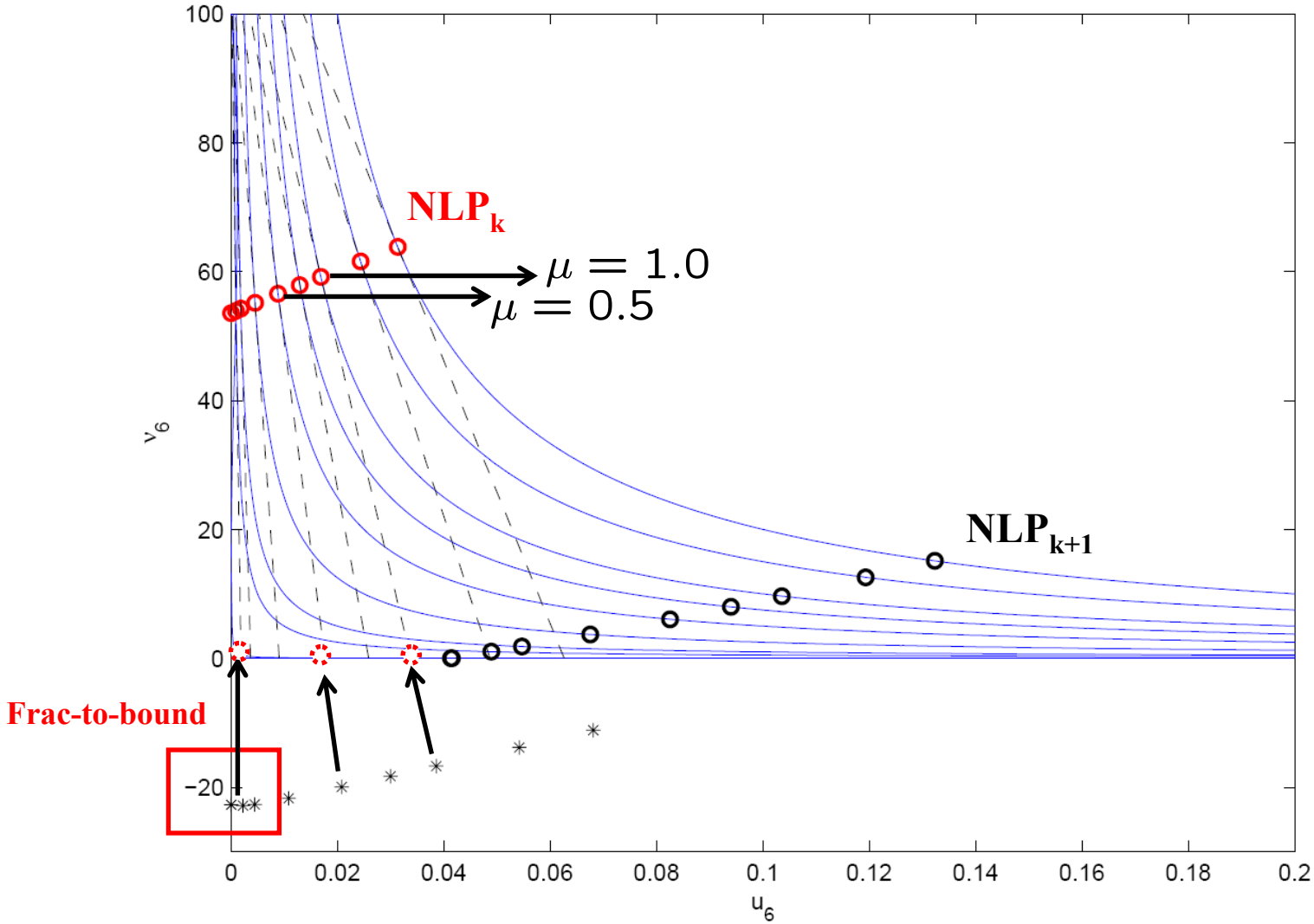
Case 2: Change in the Active-Set



Nonlinear Model Predictive Control

Case 2: Change in the Active-Set

NLP Sensitivity Provides an Infeasible Warm-Starting Point

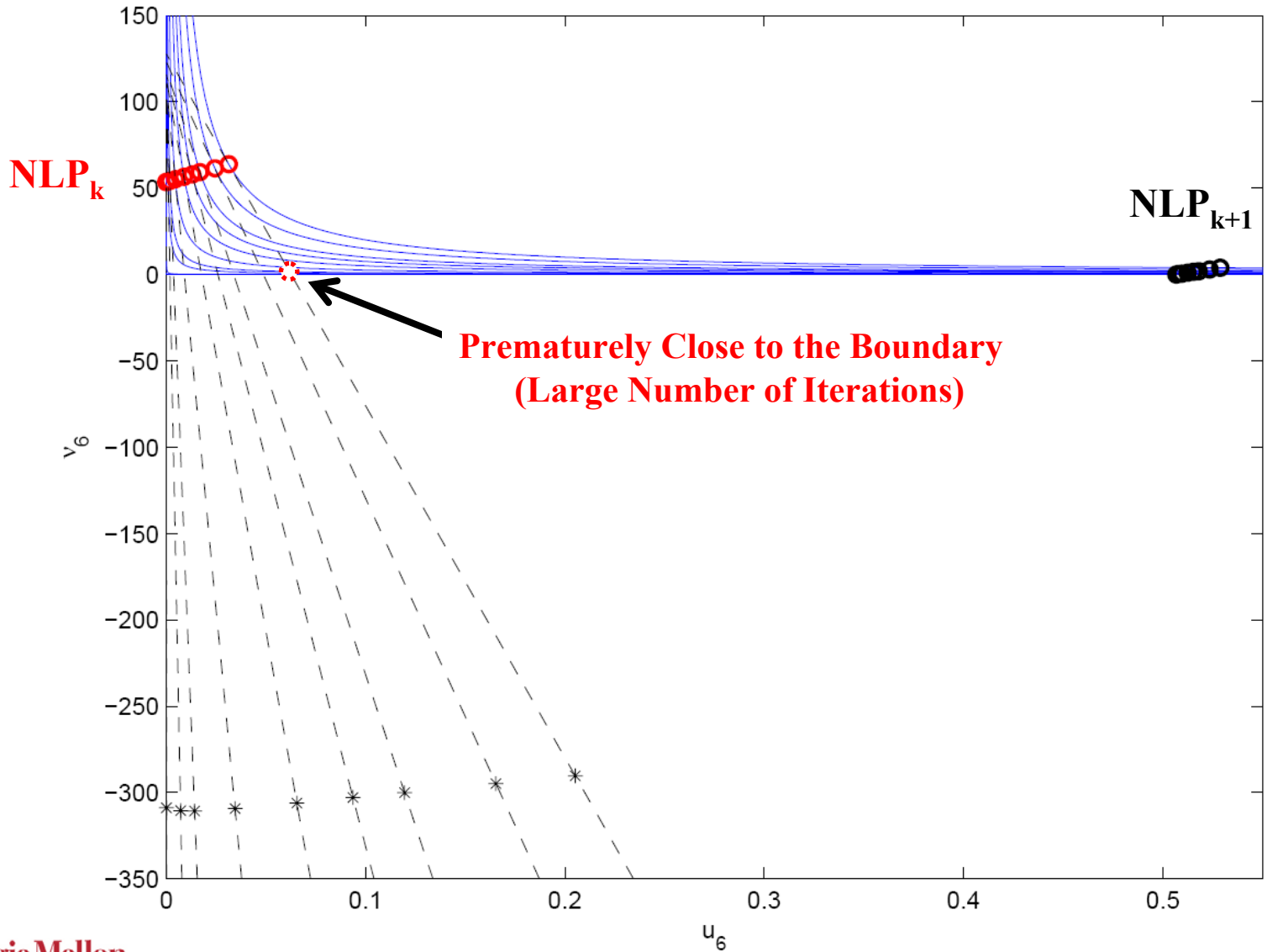


Open Questions

Open Questions

Case 3: Change in the Active-Set -Stronger Perturbations-

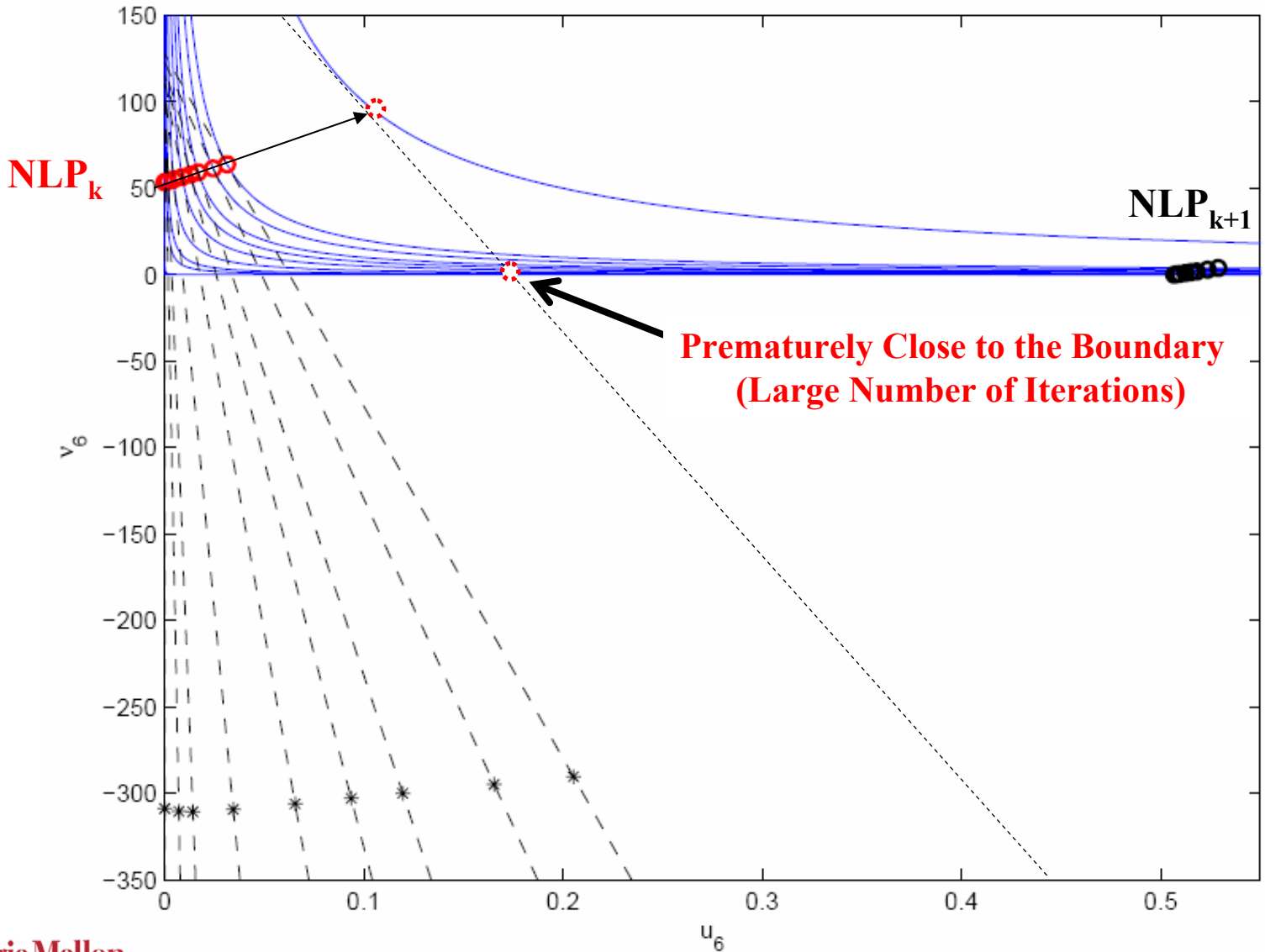
Fraction-to-the-Boundary – Fast Detection of Change



Open Questions

Case 3: Change in the Active-Set -Stronger Perturbations-

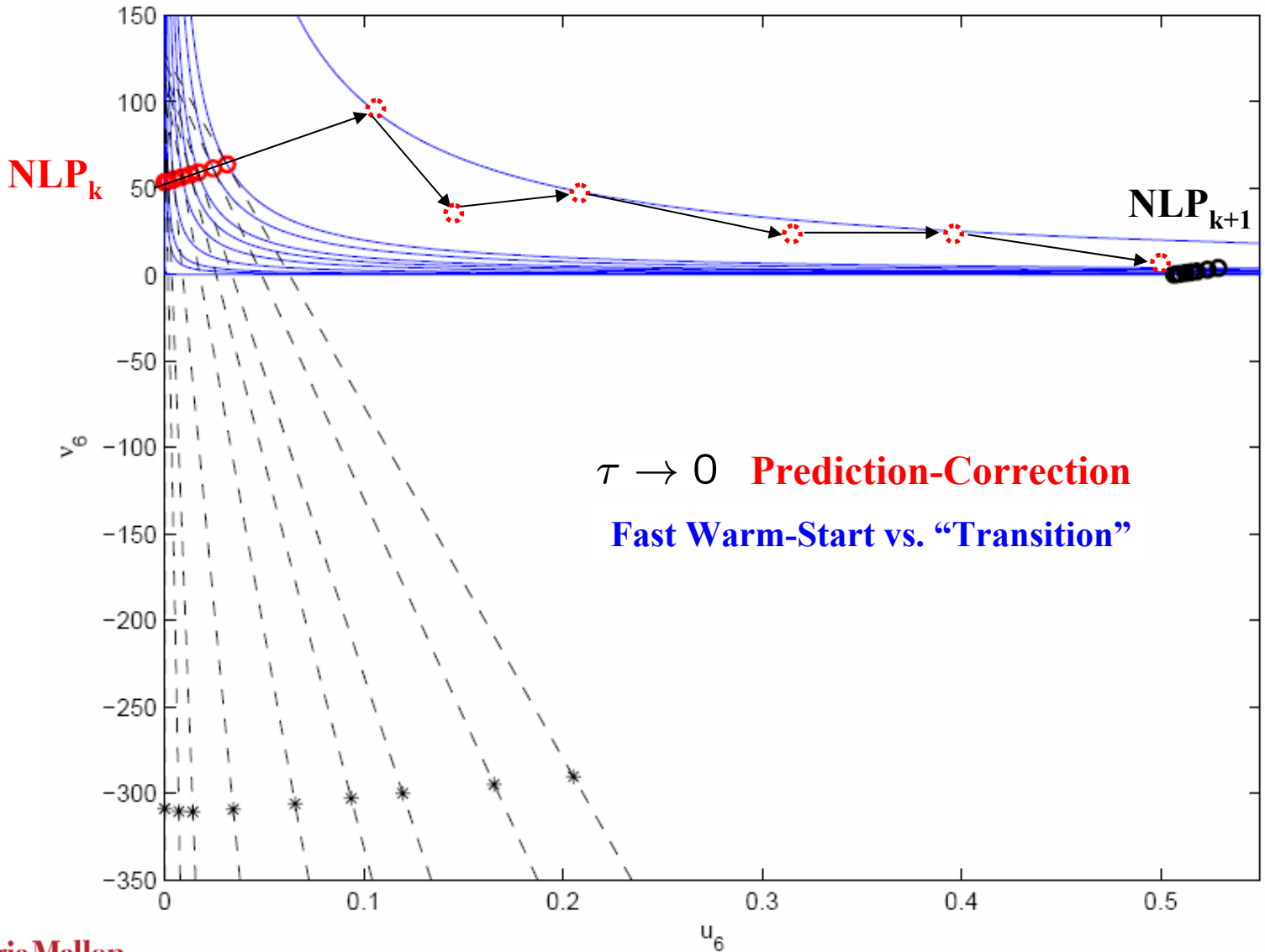
Extrapolate to large value of μ ? *-Nash and Sofer, 1996-*



Open Questions

Case 3: Change in the Active-Set -Stronger Perturbations-

Parameterize Perturbation? $p_\ell = \tau\, p_{k+1} + (1-\tau) p_k$



Summary and Conclusions

Warm-Start Crucial in Many Applications

- Active-Set vs. IPM - Active-Set Changes
- IPMs Required for Large Applications

Current Warm-Start Strategies for IPM

- Complexity Results Established for LP *-Wright 2001, Gondzio 2002-*
- Practice –Store Trial Points- Reduce Iterations but CPU Times Similar

Explore New Ideas

- Parameterize Perturbation? - Predictor-Corrector
- Strong Perturbations
  - Very Large Values of $\mu \approx O(10^1)$?
  - Extrapolation

Questions