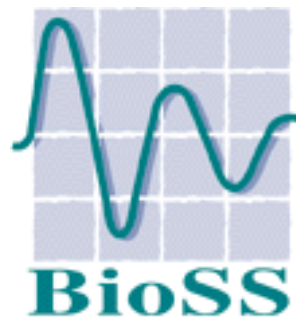


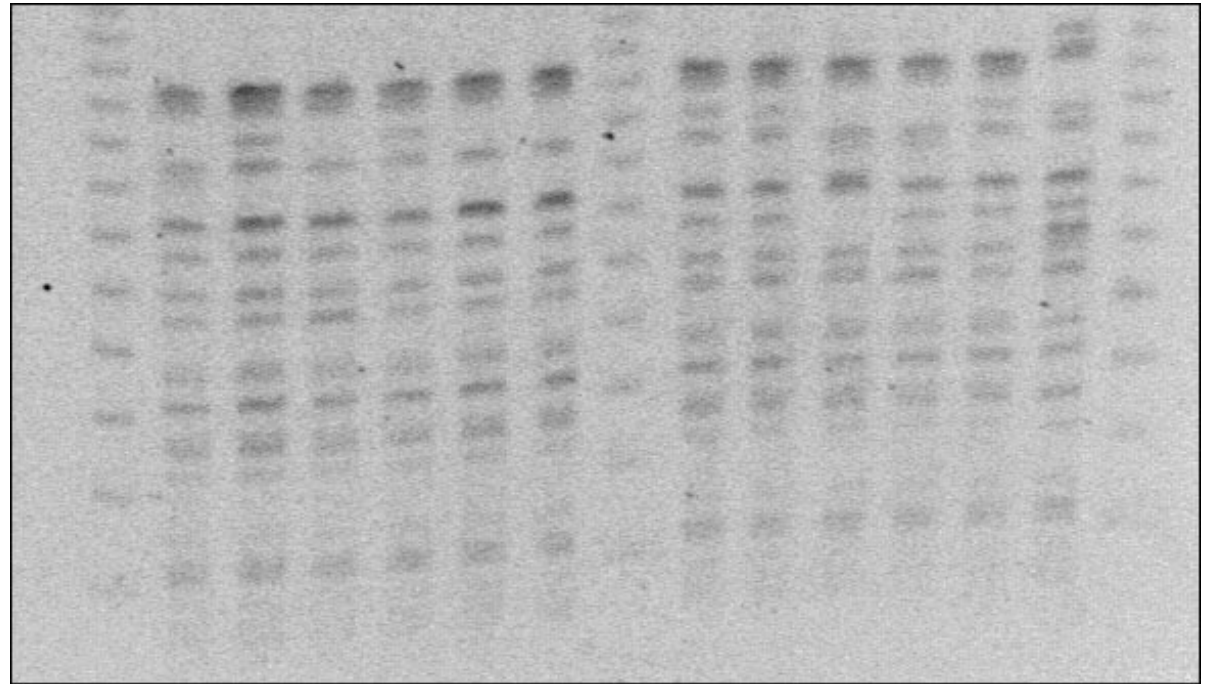
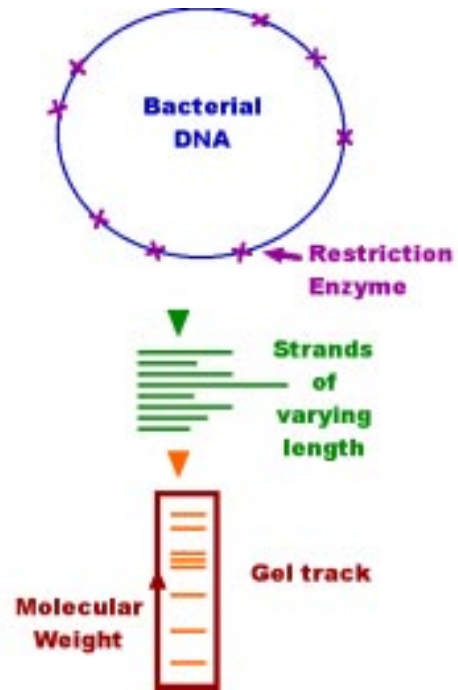
WARPING OF ELECTROPHORESIS GELS USING GENERALISATIONS OF DYNAMIC PROGRAMMING

Chris Glasbey

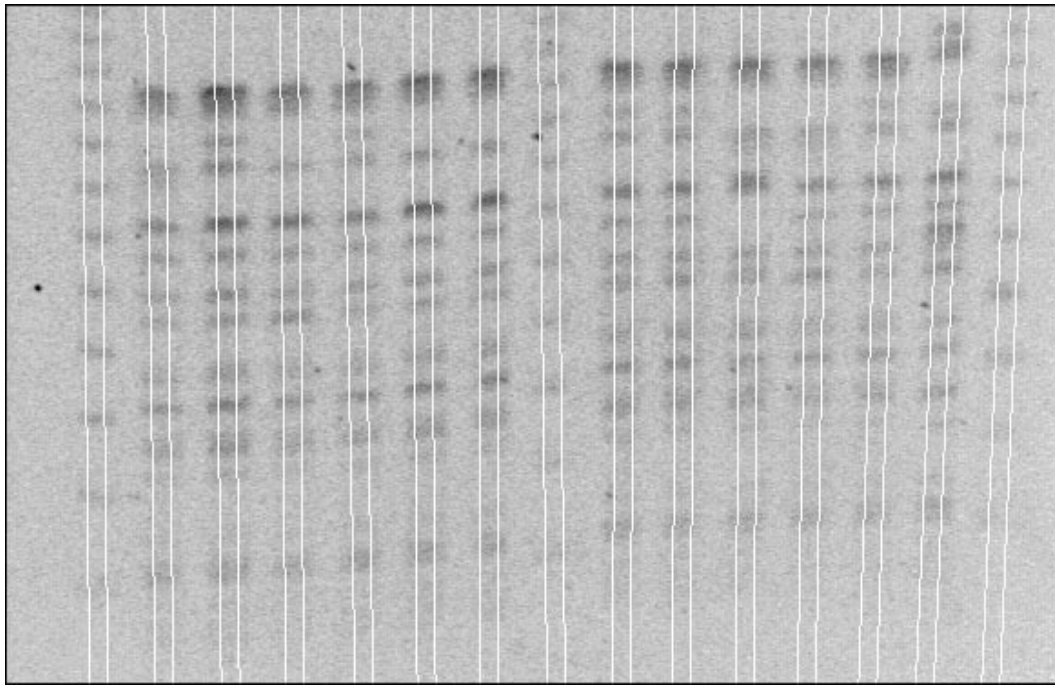


Biomathematics and Statistics Scotland

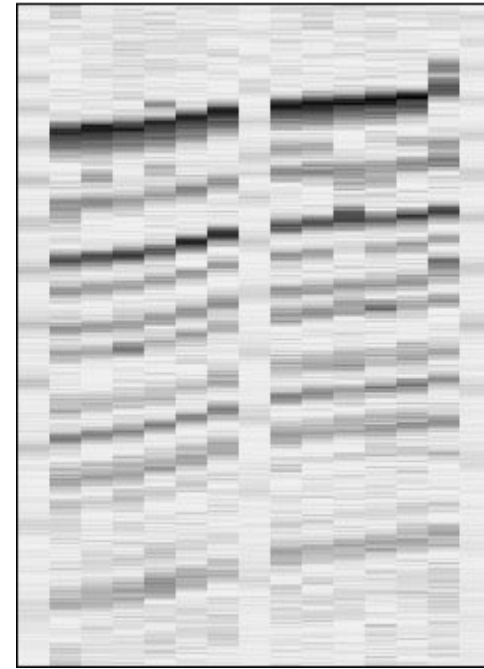
Pulsed-field gel electrophoresis (PFGE):



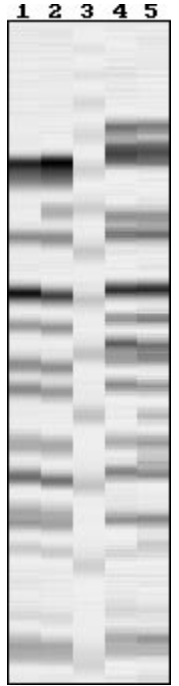
Are these new strains of *E. coli* O157?



Track markers



Y



μ

We seek the warping function (f) and lane labels (l) that minimise

$$C(f, l) = \sum_i \sum_j \left[(Y_{ij} - \mu_{i+f_{ij}, l_j})^2 + \lambda_1 (f_{ij} - f_{i-1, j})^2 + \lambda_2 (f_{ij} - f_{i, j-1})^2 \right]$$

2. Three Generalisations of Dynamic Programming (DP)

A) Iterated DP (Leung et al, 2004)

0) Initialise $f = 0$

1) For each column j in turn, use DP to minimise $C(f, l)$ w.r.t. $f_{.j}$ and l_j , given current values of all other f 's and l 's, i.e. minimise

$$\sum_i \left[\left(Y_{ij} - \mu_{i+f_{ij}, l_j} \right)^2 + \lambda_1 (f_{ij} - f_{i-1,j})^2 + \lambda_2 \left\{ (f_{ij} - f_{i,j-1})^2 + (f_{ij} - f_{i,j+1})^2 \right\} \right]$$

2) Repeat (1) until convergence

Note: This is a block-updating version of Iterated Conditional Modes (ICM – Besag, 1986)

B) Stochastic IDP

0) Initialise $f = 0$ and $T = 10^5$

1) Simulate $e_{ijfl} \sim U[0, T]$ for all i, j, f, l

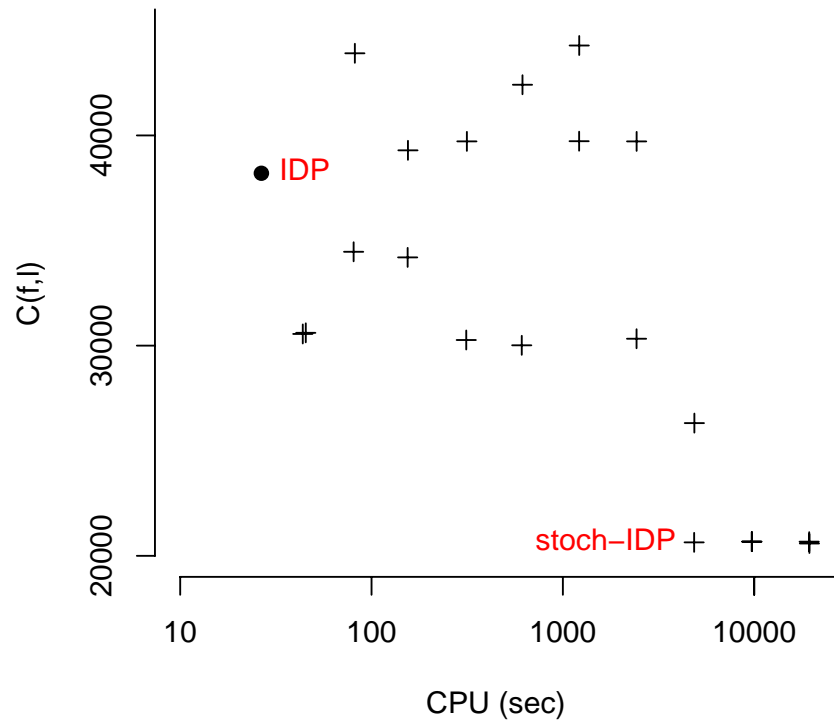
Then, for each column j in turn, use DP to minimise

$$\sum_i \left[\left(Y_{ij} - \mu_{i+f_{ij}, l_j} \right)^2 + e_{ij, f_{ij}, l_j} + \lambda_1 (f_{ij} - f_{i-1, j})^2 + \lambda_2 \left\{ (f_{ij} - f_{i, j-1})^2 + (f_{ij} - f_{i, j+1})^2 \right\} \right]$$

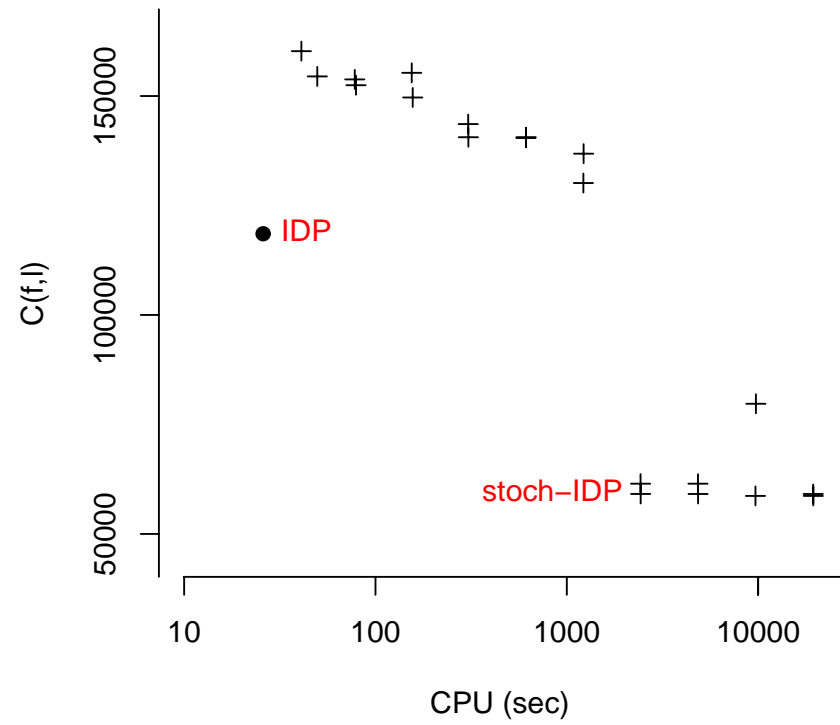
w.r.t. $f_{.j}$ and l_j

2) Reduce $T \searrow \alpha T$, and repeat (1) until f unchanged, with α such that $\{\# \text{ iterations}\} \simeq 2^6, 2^7 \dots$

Gel 1



Gel 2



C) Forward-Backwards sampler + simulated annealing

0) Initialise $f = 0$ and $T = 10^5$

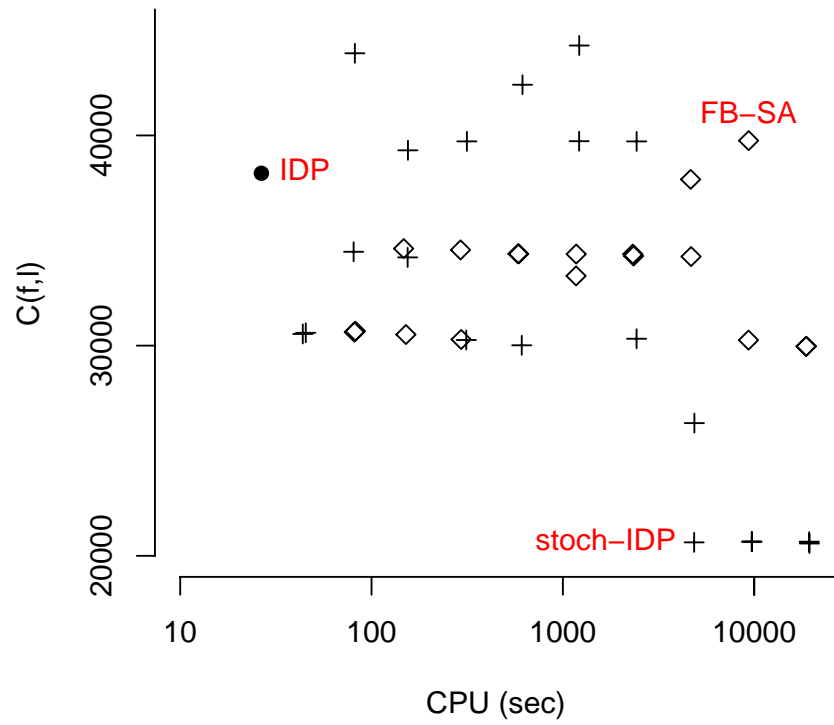
1) For each column j in turn, use Forward-Backwards Gibbs Sampler (FB) (Scott, 2002) to sample $f_{.j}$ and l_j with probability $\propto e^{-C(f,l)/T}$, given current values of all other f 's and l 's, i.e. sample from

$$\propto \exp \left[-\frac{1}{T} \sum_i \left((Y_{ij} - \mu_{i+f_{ij}, l_j})^2 + \lambda_1 (f_{ij} - f_{i-1,j})^2 + \lambda_2 \{ (f_{ij} - f_{i,j-1})^2 + (f_{ij} - f_{i,j+1})^2 \} \right) \right]$$

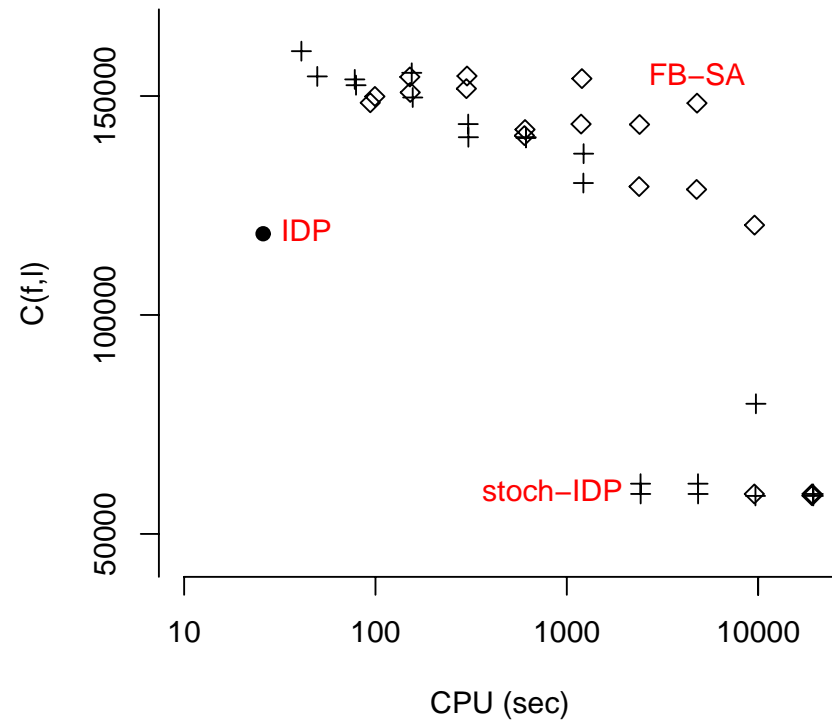
2) Reduce $T \searrow \alpha T$, and repeat (1) until f unchanged

Note: This is a form of block-updating Gibbs sampler (Roberts and Sahu, 1997)

Gel 1



Gel 2



Finally, a modification to IDP

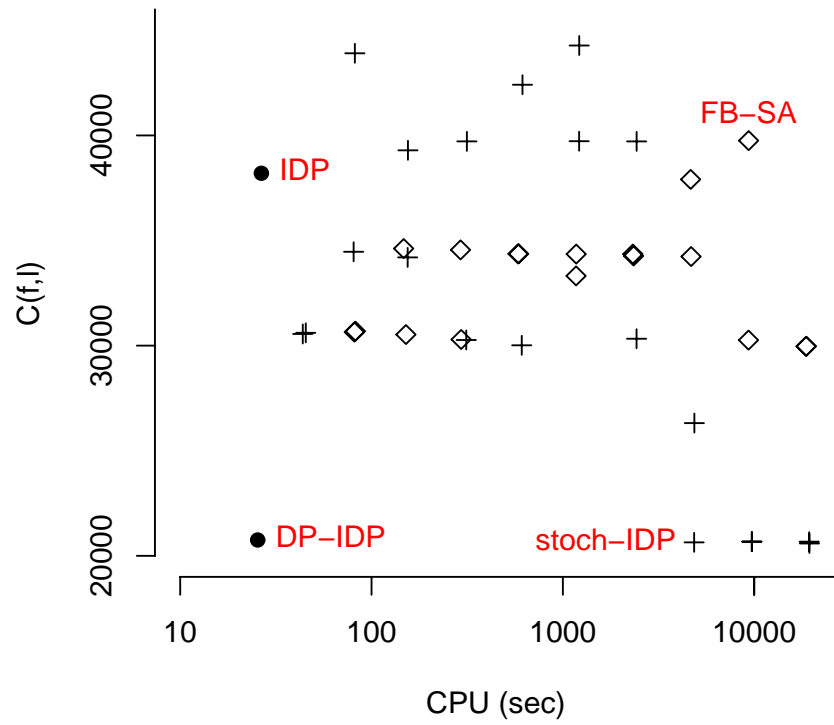
replace step (0) by

0) For each column j in turn, use DP to minimise

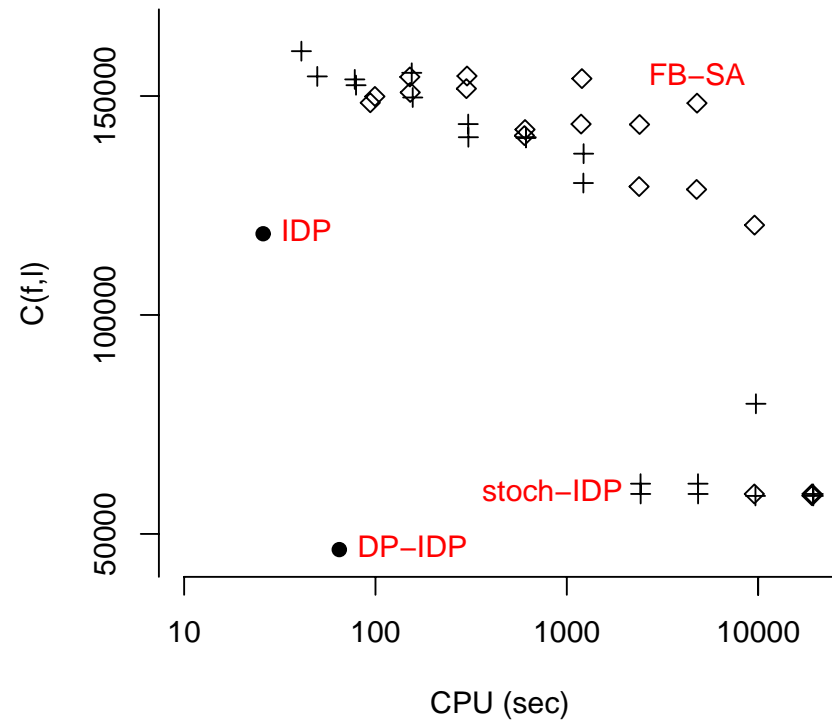
$$\sum_i \left[\left(Y_{ij} - \mu_{i+f_{ij}, l_j} \right)^2 + \lambda_1 (f_{ij} - f_{i-1, j})^2 \right]$$

w.r.t. $f_{.j}$ and l_j

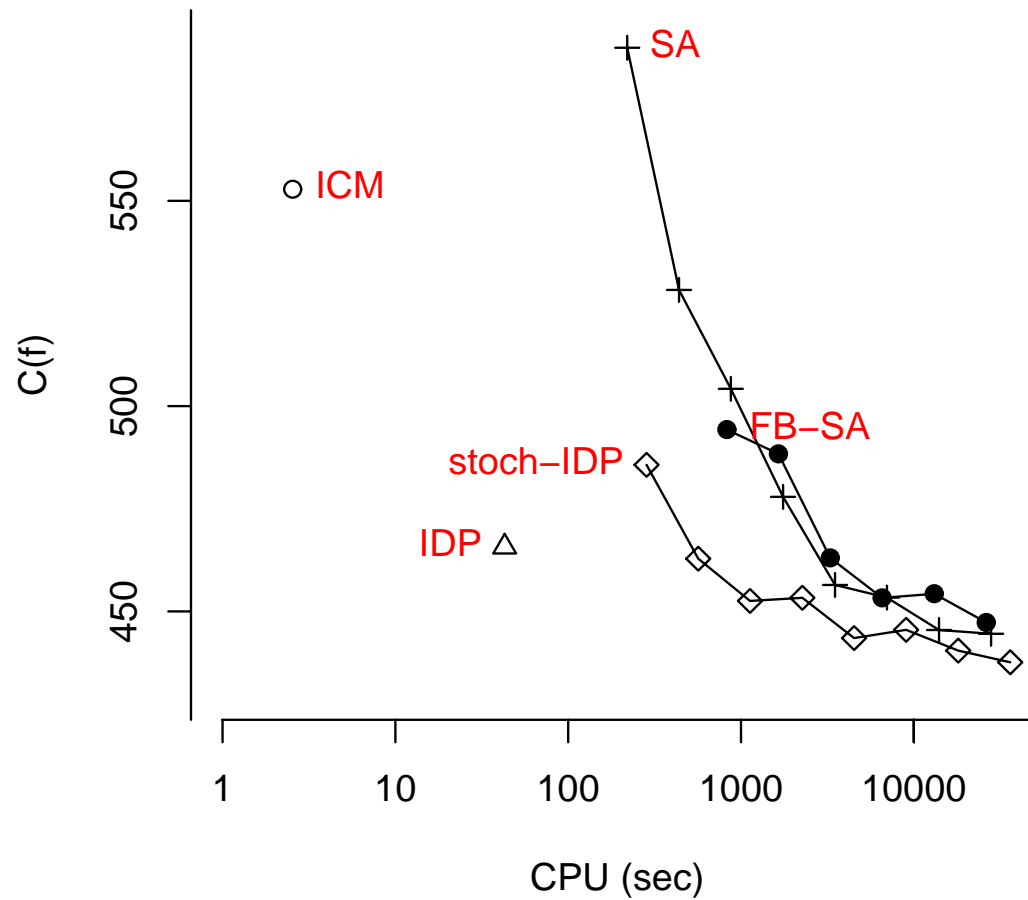
Gel 1



Gel 2



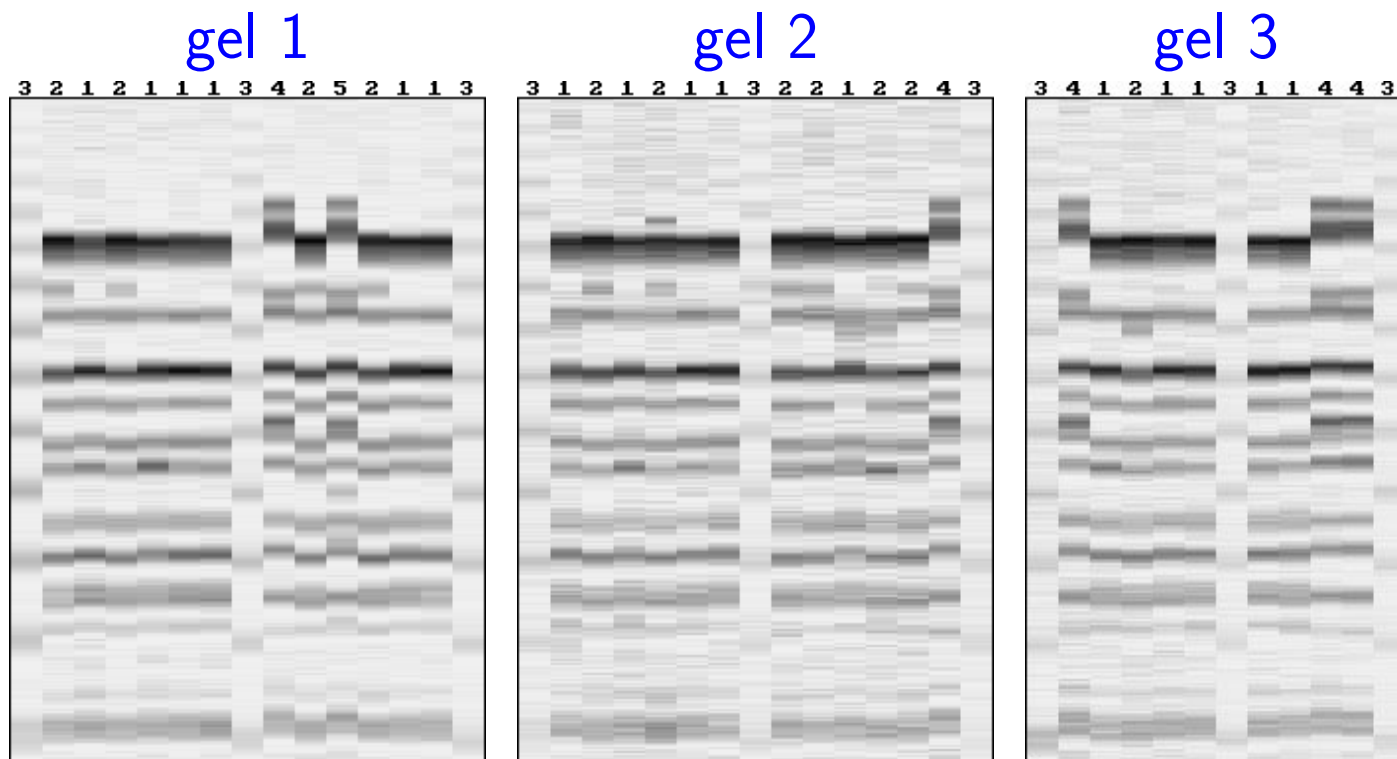
In contrast, in an image restoration application, stochastic IDP was best:



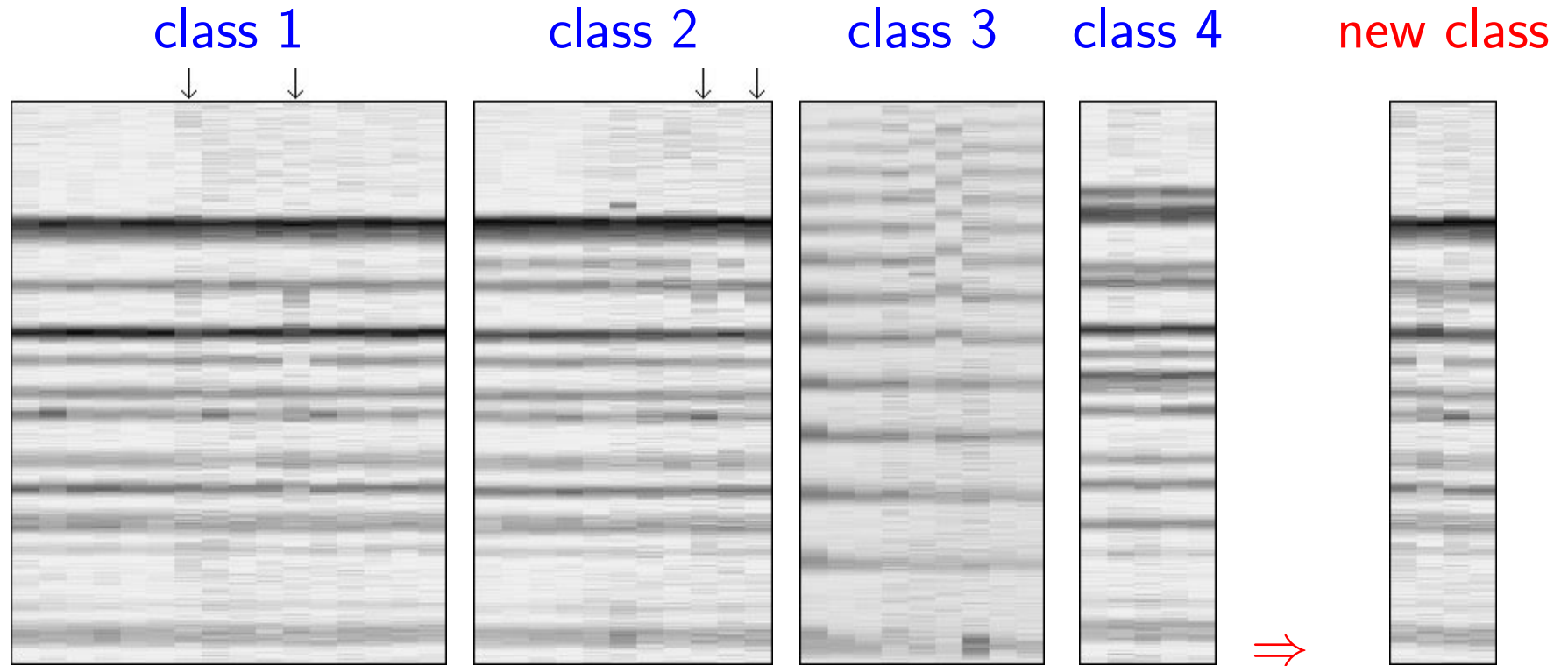
3. Results

PFGE

$Y_{f_{ij}^{-1}, j}$, where f^{-1} is an inverse function in the first index, defined such that $f_{f_{ij}^{-1}, j} \equiv i$.

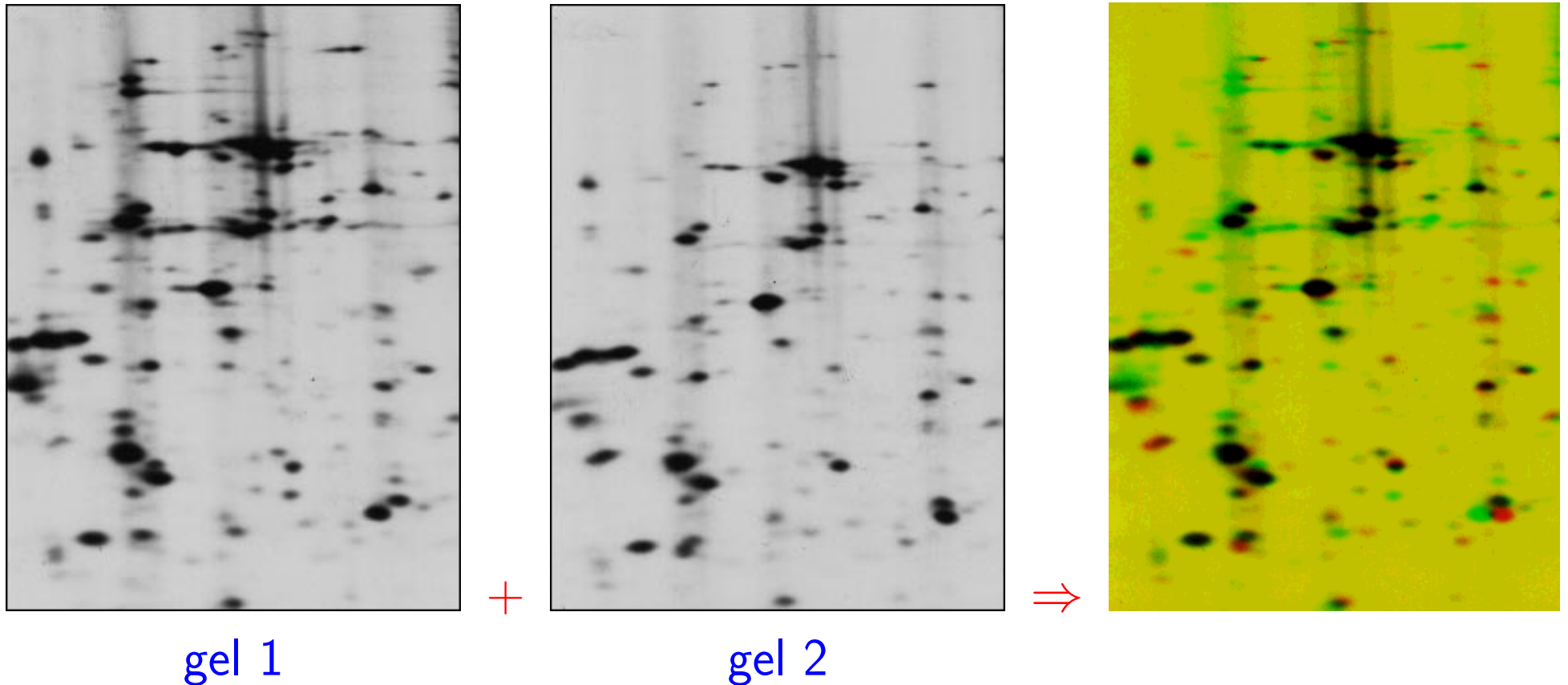


Re-ordered



Arrowed columns visually identified as belonging to *E. coli* strains not in database

The methods can also be used to align 2-D gels:



Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of two strains of malarial parasite

4. Summary

Dynamic programming (DP) is a fast, elegant method for finding the global solution to a class of optimisation problems, including alignment of single lanes in PFGE gels.

However, it cannot be used more generally for image warping.

We considered 3 generalisations of DP:

- Iterated DP (with different starts)
- Stochastic IDP
- Forward-Backwards sampler (FB) + simulated annealing (SA)

In the PFGE application, DP-IDP performed best.